



Bachelorarbeit in Informatik: Games Engineering

# **Klassifizierung chronischer Wunden durch Machine Learning**

**Luca Mario Hohmann**





Bachelorarbeit in Informatik: Games Engineering

## **Klassifizierung chronischer Wunden durch Machine Learning**

## **Classification of chronic wounds with machine learning**

Autor: Luca Mario Hohmann  
Aufgabensteller: Prof. Dr. Helmut Krcmar  
Betreuer: M.Sc. Kai Klinker  
Abgabedatum: 15.07.2020



Ich versichere, dass ich diese Bachelorarbeit in Informatik: Games Engineering selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

Herrsching,

Luca Mario Hohmann

## Danksagungen

Ich danke meiner Familie, die mich immer auf meinem Weg unterstützt. Außerdem möchte ich gerne meinem Betreuer Kai Klinker für die gemeinsame Arbeit am Lehrstuhl der Wirtschaftsinformatik der TUM und der anschließenden Umsetzung der Arbeit in diese Bachelorarbeit danken.

# Zusammenfassung

Um der täglich großen Zahl an Wundpatienten in Deutschland besser helfen zu können, benötigen Pflegekräfte effiziente und einheitliche Versorgungsprozesse und -standards. Um Pflegekräfte in ihrer Arbeit mit Informationssystemen zu unterstützen, werden neue Hardware und Software basierte Lösungen benötigt. Daher wurden in dieser Bachelorarbeit mithilfe der Design Science Methode zwei aktuelle Open Source Machine Learning Modelle jeweils zweimal trainiert, um chronische Wunden in Bildern lokalisieren und deren Wundtyp bestimmen zu können. Dafür wurden die drei häufig vorkommenden Wundtypen Dekubitus, diabetisches Fußsyndrom und venös arterielle Ulzera ausgewählt. Die beiden Machine Learning Modelle erreichten nach dem ersten Training eine mAP von 23.4% und 22.2% sowie F1-Werte von 0.356 und 0.295. Beim zweiten Training wurden mAPs von 20.15% und 21.88%, sowie F1-Werte von 0.269 und 0.276 erreicht. Da für das Training der Machine Learning Modelle nur wenig öffentlich zugängliche, medizinisch verifizierte Wundbilder vorhanden waren, wurde ebenfalls eine webbasierte Annotationsplattform mit Fokus auf den täglichen Einsatz bei Pflegepersonal entwickelt. Diese können durch wenige Annotationen am Tag in großen Mengen ihr Wissen nutzen, um eine medizinisch korrekte Datenbank mit annotierten Wundbildern für Machine Learning zu erzeugen. Bei den nicht medizinischen Testpersonen der Webseite wurde beim System-Gebrauchstauglichkeit Test ein leicht überdurchschnittlicher Gesamtwert von 71 erreicht.

# **Abstract**

In order to help the high number of wound patients in Germany, healthcare professionals need efficient and universally coherent treatment processes. To improve healthcare professionals work with information systems, new hardware and software-based solutions must be implemented. Hence this bachelor's thesis trained two current open source machine learning models, each two times, in order to localize chronic wounds in images and determine their wound type. This thesis focused on three of the most frequent wound types, Pressure Ulcer, Diabetic Foot Ulcer and Venous-Arterial Ulcer. The machine learning models reached mAPs of 23.4% and 22.2% and F1-Scores of 0.356 and 0.295 after the first training. Following the second training, the models reached mAPs of 20.15% and 21.88% as well as F1-Scores of 0.269 and 0.276. For the training of machine learning models in wound detection, only few medically verified data sources were available. Therefore, this thesis implemented a web-based annotation platform for daily usage by healthcare professionals. With only a few annotations a day, healthcare workers may contribute to a medically accurate database of wound images and annotations for machine learning. The non-medical testers of the website rated the tool with a System-Usability-Scale result of 71, which is slightly above the average result.

# Inhaltsverzeichnis

<b>Danksagungen</b>	iii
<b>Zusammenfassung</b>	iv
<b>Abstract</b>	v
<b>1 Einleitung</b>	1
1.1 Problemstellung . . . . .	1
1.2 Zielsetzung . . . . .	2
<b>2 Methodisches Vorgehen</b>	3
<b>3 Rigor</b>	5
3.1 Wunden . . . . .	5
3.1.1 Wundgewebe . . . . .	5
3.1.2 Heilungsprozess . . . . .	6
3.1.3 Chronische Wunden . . . . .	7
3.2 Maschinelles Lernen . . . . .	12
3.2.1 Grundlegende Begriffe . . . . .	12
3.2.2 Grundlegende Algorithmen . . . . .	18
3.2.3 Verwendeter Algorithmus . . . . .	22
3.3 Weitere Einsatzmöglichkeiten von maschinellem Lernen in der Wundmedizin	24
3.4 Weitere technische Hilfsmittel in der Wundpflege . . . . .	26
<b>4 Relevance</b>	28
4.1 Aktueller Wundpflegeprozess . . . . .	28
4.2 Vergleich der Prozesse . . . . .	29
4.3 Potenzial und Verbesserungen durch das Programm . . . . .	34
<b>5 Design Cycles</b>	35
5.1 Design Cycle 1 . . . . .	35
5.1.1 Grundlagen und Implementierung . . . . .	35
5.1.2 Nutzertests . . . . .	38
5.1.3 Resultate . . . . .	38
5.1.4 Gelerntes . . . . .	39
5.2 Design Cycle 2 . . . . .	39
5.2.1 Grundlagen und Implementierung . . . . .	39

5.2.2	Nutzertests . . . . .	41
5.2.3	Resultate . . . . .	42
5.2.4	Gelerntes . . . . .	42
5.3	Design Cycle 3 . . . . .	43
5.3.1	Grundlagen und Implementierung . . . . .	43
5.3.2	Nutzertests . . . . .	43
5.3.3	Resultate . . . . .	44
5.3.4	Gelerntes . . . . .	44
5.4	Design Cycle 4 . . . . .	44
5.4.1	Grundlagen und Implementierung . . . . .	44
5.4.2	Tests . . . . .	47
5.4.3	Resultate . . . . .	48
5.4.4	Gelerntes . . . . .	53
5.5	Design Cycle 5 . . . . .	53
5.5.1	Grundlagen und Implementierung . . . . .	53
5.5.2	Tests . . . . .	55
5.5.3	Resultate . . . . .	55
5.5.4	Gelerntes . . . . .	58
<b>6</b>	<b>Implikationen</b>	<b>61</b>
6.1	Theoretische Implikationen . . . . .	61
6.2	Praktische Implikationen . . . . .	61
<b>7</b>	<b>Zukünftige Forschungsfragen</b>	<b>63</b>
<b>8</b>	<b>Abschließende Zusammenfassung</b>	<b>64</b>
	<b>Abbildungsverzeichnis</b>	<b>65</b>
	<b>Tabellenverzeichnis</b>	<b>67</b>
	<b>Glossar</b>	<b>69</b>
	<b>Akronyme</b>	<b>70</b>
	<b>Literatur</b>	<b>71</b>

# 1 Einleitung

## 1.1 Problemstellung

In Deutschland wurden im Jahr 2012 insgesamt 2.7 Millionen Patienten mit Wunden behandelt [1]. Von diesen Patienten musste ein Drittel länger als 8 Wochen behandelt werden, was zu einer Einstufung als Patient mit chronischen Wunden führt. Besonders Wundpatienten mit Dekubitus, diabetischem Fußsyndrom und Ulcus Cruris sind anfällig für einen chronischen Wundverlauf [1] [2]. Bei chronischen Wunden müssen Patienten über viele Wochen medizinisch begleitet werden und regelmäßig die Wunde dokumentieren lassen [3]. Dies ist ein aufwändiger Prozess, der von unterschiedlichem Personal in unterschiedlichen Einrichtungen durchgeführt werden kann. Dadurch sind unterschiedliche Bemessungsgrundlagen für die Dokumentation und Bewertung der Wundheilung durch unterschiedliche Vorgaben und Erfahrung des Personals möglich. Besonders mit Blick auf eine Umfrage des Instituts für Gesundheits-System Forschung bezüglich der Versorgung chronischer Wundpatienten in Deutschland, zeigt sich die Verbreitung der Behandlung von chronischen Wunden unter deutschen Ärzten [4]. Bis zu 98.7% der 853 befragten Ärzten gaben an, Patienten mit chronischen Wunden zu behandeln. Damit wird in nahezu allen Praxen Wissen über die Dokumentation und die Behandlung von chronischen Wunden, sowohl von den Ärzten als auch dem medizinischen Hilfspersonal, benötigt. Neben der hohen Zahl von Ärzten, die mit der Behandlung chronischer Wunden in Kontakt kommen, sind auch andere, nicht medizinische Berufe angehalten, sich um die akkurate Wundbehandlung zu kümmern. Denn nach dem Versorgungsstrukturgesetz aus dem Jahre 2015, ist es erlaubt diese eigentlich ärztliche Tätigkeit auch an nichtärztliche Berufe zu delegieren [5]. Dadurch sollen ärztlich unversorgte Regionen [6] durch anderen Personalien, wie zum Beispiel medizinische Fachangestellte oder Pflegepersonal, unterstützt werden. Damit eine nichtärztliche Person eine solche Aufgabe übernehmen darf, muss sie entsprechende Qualifikationen vorweisen können [2]. Allerdings leiden diese Berufsgruppen auch unter Fachkräftemangel, wie eine KOFA Studie aus dem Jahr 2019 zeigt [7]. Dadurch entsteht ein hoher Druck auf den verbleibenden Fachkräften, der in Kombination mit unterschiedlichen Ausbildungen zu unterschiedlichen Ergebnisse und auch Dokumentationslücken führen kann [2]. Dies wiederum kann zu Problemen beim der zukünftigen Behandlung der Wunde führen und den Patienten weiter belasten. Damit die wenigen Fachkräfte eine höhere Effizienz während der Wundpflege erreichen können und somit dadurch entlastet werden, müssen digitale Lösungen erreicht werden [8].

## 1.2 Zielsetzung

Im Bereich des Maschinelles Lernen (ML) gibt es bereits in anderen medizinischen Bereichen vielversprechende Ergebnisse. Besonders mit Blick auf Brustkrebs- und Alzheimererkennung werden große Hoffnungen in diese Technologien gesetzt [9]. Um den Prozess der Wundbehandlung zu erleichtern und eine deutlichere Vereinheitlichung bei der Bewertung von Wunden voranzubringen, wird in dieser Arbeit ein ML basiertes Wunderkennungs- und Klassifizierungssystem für vier verschiedenen Wundtypen, wobei zwei Typen in eine Klasse zusammengefasst werden, implementiert. Dieses soll dabei helfen, vier häufig vorkommende Wundtypen in Bildern zu erkennen und korrekt einordnen zu können. Besonderer Fokus wird daraufgelegt, den Stand von Open Source ML Technologie im medizinischen Bereich genauer zu erforschen. Des Weiteren soll ein webbasiertes Annotationsprogramm für die Generierung großer, medizinisch korrekte Datensätze für ML entwickelt werden. Dadurch soll es in Zukunft möglich sein, mit der Hilfe von medizinischem Personal korrekte Annotationen von Wundbildern durchzuführen und anschließend beliebige ML Modelle damit trainieren zu können. Denn zum aktuellen Zeitpunkt gibt es nur wenige öffentliche Klassifizierungen von Wundbildern, was eine Implementierung eines ML System deutlich erschwert.

Für die Erforschung beider Systeme wird zunächst die eingesetzte Design Science Methode beschrieben. Im Anschluss werden Grundlagen in den bearbeiteten Bereichen vermittelt, um anschließend aktuelle Arbeiten im Bereich der ML gestützten Wundpflege besser analysieren zu können. Danach werden die Anforderungen an Programme in der Pflege mithilfe von Analysen vorheriger Arbeiten in diesem Bereich identifiziert. Anschließend wird das webbasierte Annotationsprogramm iterativ implementiert und mit Probanden getestet. Anschließend werden durch das Tensorflow Framework [10] aktuelle ML Technologien und Techniken implementiert, um die automatische Erkennung und Klassifizierung von Wunden zu untersuchen. Nach der Auswertung und dem Vergleich der Ergebnisse zu anderen ML gestützten Systemen, werden die Implikationen der Ergebnisse erläutert. Danach schließen ein Ausblick und eine Zusammenfassung der vorherigen Kapitel die Bachelorarbeit ab.

## 2 Methodisches Vorgehen

Der in dieser Bachelorarbeit verfolgte Arbeitsprozess basiert auf der Design Science Methodik. Diese ist, wie der Name andeutet, besonders auf das Design des zu erforschenden Produkts fokussiert und versucht die Bedürfnisse der intendierten Nutzer maximal in den Arbeitsprozess einzubinden [11]. Die Design Science Methodik baut ihr Wissen auf drei grundlegenden Teilen auf. Im Fokus der Arbeit stehen die Design Cycles, welche zwischen schneller Implementation und Evaluation wechseln, um so ein Artefakt im Bereich der Informationssysteme zu erhalten [12]. Um die Design Cycles maximal unterstützen zu können, wird ein Relevance Cycle benötigt. In diesem Abschnitt werden wichtige Anforderungen an das Artefakt und dessen Einsatz im intendierten Umfeld gesammelt. Dies geschieht mithilfe von Ethnographien, Interviews, Fokusgruppen und der Analyse des zu lösenden Problems. Denn das grundlegende Problem muss klar und im Detail verstanden sein, um gezielt auf eine Lösung hinzuarbeiten zu können. Neben dem Relevance Cycle wird ein Rigor Cycle durchgeführt, in welchem das aggregierte Wissen aus wissenschaftlichen Arbeiten genutzt wird, um den aktuellen Stand im untersuchten Bereich festzuhalten [11]. Zudem können so grundlegende Design Theorien identifiziert werden, um die Evaluation der Artefakte in den Design Cycles an diesen zu orientieren [11].

Die einzelnen Schritte dieser Arbeit sind in Abbildung 2.1 dargestellt. Anhand der identifizierten Schritte, wurde zunächst ein Rigor Cycle durchgeführt. In diesem ist Literatur aus den behandelten Bereichen untersucht worden. Diese umfassen den Bereich der Wundpflege, sowie ML und aktuelle Assistenzsysteme in der Wundpflege. Im Anschluss wurde ein Relevance Cycle durchgeführt, welcher im Gegensatz zum normalen Vorgehen aus einer Analyse von vorangegangenen Design Science Arbeiten in diesem Bereich besteht. Denn die zum Zeitpunkt des Relevance Cycles vorherrschende SARS-Cov2 Situation stellte einerseits ein hohes Infektionsrisiko für die potenziellen Teilnehmer, sowie eine zusätzliche Belastung des Pflegepersonals während der Epidemie dar. Auf die Rigor und Relevance Cycles aufbauend, wurden fünf Design Cycles angesetzt. Diese sind in eine Gruppe von drei Design Cycle für die Annotationsplattform, sowie eine zweite Gruppe von zwei Design Cycles für das ML System aufgeteilt worden. Diese Aufteilung soll eine klare Struktur und einen vollständigen Fokus auf die Optimierung des zu untersuchenden Systems mit sich bringen. Durch diese Aufteilung sind auch die unterschiedlichen Testmethoden leicht zu separieren und detaillierter beschreibbar. Denn die ersten drei Design Cycles erfordern Probanden, die mit dem Artefakt in Berührung kommen und dieses nutzen, während in den letzten beiden Design Cycles keine Probanden gebraucht werden, da das ML System automatisch auf eingespeiste Bilder reagiert und damit keine Bedienmöglichkeit durch Probanden vorhanden ist. Die Benutzertests sind ein essentieller Bestandteil jedes Design Cycles. Das zuvor implementierte

## 2 Methodisches Vorgehen

---

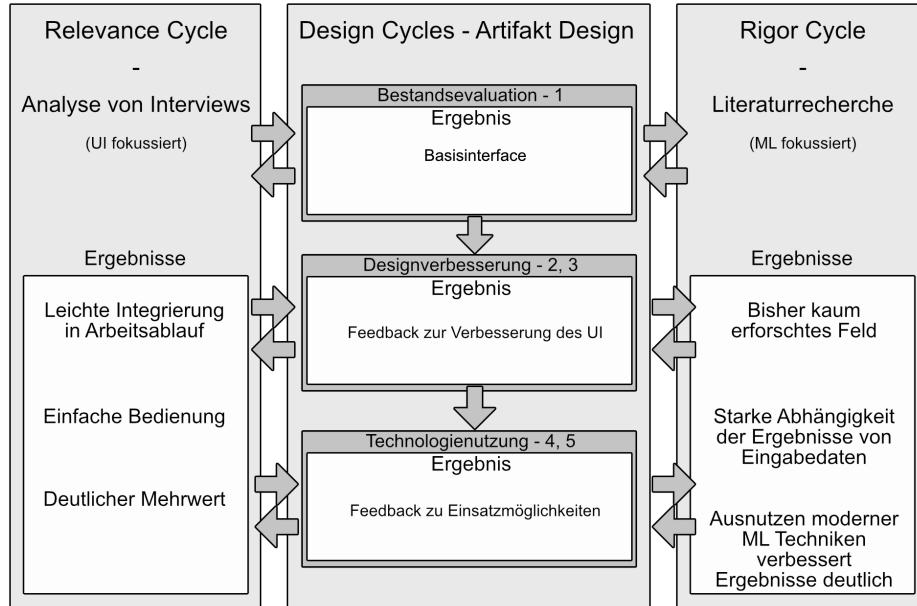


Abbildung 2.1: Darstellung der einzelnen Cycles der Design Science Methode in dieser Arbeit

Artefakt wird dabei mit einer adäquaten Testmethode von den ausgewählten Nutzern getestet. Da die Design Science Methode stark design orientiert arbeitet, wurde die 'Think a Loud' Methode für alle Tests ausgewählt. Dabei werden die Nutzer gebeten einige Aufgaben zu erfüllen und dabei alles laut zu Beschreiben was Sie lesen, denken, machen möchten und auch fühlen [13]. Während der Ausführung der Aufgaben macht der Testleiter Notizen, um die Vorgehensweise und die Aussagen des Probanden zu dokumentieren. In den ersten Design Cycles wurde das ML Modell deaktiviert und eine Erkennung von Objekten mittels zufallsbasierten Begrenzungsboxen simuliert. Des Weiteren wurden in diesen Design Cycles keine Wundbilder verwendet, so dass kein Fachwissen benötigt wird, um die Aufgaben zu lösen. Die Testgruppe besteht dabei aus sechs Teilnehmern, die Bekannte von mir sind. Die Gruppengröße wurde basierend auf einem Artikel von Jakob Nielson [14] ausgewählt, um möglichst viele, nicht repetitive Informationen für ein Programm von kleiner Größe zu erhalten. Vier Teilnehmer studieren im Bereich Informatik oder sind im Bereich IT Administration tätig. Zwei der Teilnehmer sind nicht im technischen Bereich tätig. Durch diese Verteilung wurde einerseits versucht eine Heterogenität von technischen Fähigkeiten zu erhalten, aber auch mit technischem Wissen Fehler im Programm zu offenbaren. Im dritten Design Cycle wurden zudem zwei weitere Personen, die im weiteren Bereich mit Wunden zu tun haben, gebeten das Programm zu testen und zu bewerten.

# **3 Rigor**

Zunächst wird der aktuelle Stand in der Wundpflege sowie die für diese Arbeit relevanten Wundtypen erläutert. Anschließend wird der Begriff des ML sowie einige bekannte Algorithmen genauer erklärt und im Anschluss der zugrundeliegende Algorithmus des verwendeten TensorFlow Modells beschrieben. Daraufhin wird auf den aktuelle Stand von anderen ML Ansätzen zur Unterstützung in der Wundpflege eingegangen. Abschließend werden weitere, nicht ausschließlich auf ML basierende, technische Hilfsmittel für die Wundpflege betrachtet.

## **3.1 Wunden**

### **3.1.1 Wundgewebe**

Das Ziel dieser Arbeit ist es den Wundpflegeprozess zu verbessern und mit ML zu unterstützen. Um den aktuellen und vorgeschlagen Prozess, welche beide in Abschnitt 4.2 genauer analysiert werden, besser zu verstehen, stellt diese Arbeit im nachfolgenden Abschnitt chronische Wunden im Allgemeinen, sowie die in dieser Arbeit behandelten Wundtypen vor. Bei Wunden gibt es verschiedene Arten von Geweben die sich je nach Typ und Stadium unterscheiden können. Granulationsgewebe stellt dabei eine wichtige Gewebeart dar. Es ist vor allem durch die rote Farbe und körnige Struktur erkennbar. Das Gewebe ist von Kapillaren, kleinen Blutgefäßen durchzogen und beginnt bei geringem Druck zu bluten. Eine weitere wichtige Gewebeart sind die epithelialen Zellen. Diese dienen als äußerste Schicht der Haut und verschließen jede Wunde. Zu Beginn ist das Gewebe leicht rosa und ebenfalls sehr empfindlich, weshalb es zunächst von Narbengewebe geschützt werden muss. Narbengewebe hat im Vergleich zu anderen Gewebearten keinerlei Flexibilität ist dadurch aber besonders gut als Schutzmechanismus geeignet. Neben Narbengeweben existiert auch nekrotisches Gewebe, welches alle abgestorbenen Zellen beschreibt. Dieses Gewebe ist oft schwarz oder braun und findet sich häufig am Rand und der Basis der Wunden. Eine weitere oft vorkommende Gewebestruktur ist Fibringewebe. Dieses ist in den meisten Fällen weißlich oder leicht gelblich. Das Gewebe entsteht durch körpereigenes Fibrin welches einerseits Blut gerinnt und andererseits eine Matrix spannen kann. Die letzte wichtige Gewebeart stellt Schorf (engl. slough) dar. Diese ist meistens gelb und muss in jedem Fall von der Wunde entfernt werden, da sie die Wundheilung stoppt, beziehungsweise verlangsamt. Dieser Schorf findet sich oft an der Wundbasis und im Wundbett verteilt. [15] [16] [17]

### 3.1.2 Heilungsprozess

Um den Ablauf der Wundheilung besser zu verstehen, werden zunächst die wichtigsten Schritte bei Wundheilungen vorgestellt. Wenn eine Person sich eine Wunde an einer Körperstelle zuzieht, beginnt der Körper mit dem Wundheilungsprozess. Dieser Ablauf besteht bei einer normal heilenden Wunde aus vier sich überlappenden Stufen. Dabei können einzelne Phasen parallel auftreten und auch eine Verlangsamung oder temporäre Umkehrung von Phasen ist je nach Patient und Situation möglich [18]. Im ersten Schritt setzt die exsudative Phase, auch inflammatorische oder reinigende Phase genannt, ein. Diese beginnt direkt nachdem sich die Verletzung zugezogen wurde und sorgt dafür dass die Fibrinwandlung stattfindet. Fibrin ist besonders an der Blutgerinnung beteiligt und sorgt so für den Blutungsstop. Im Anschluss werden die Gefäße um die Wunde herum erweitert um eine bessere Verbindung für benötigte Stoffe zu ermöglichen. Zu den ersten Stoffen, die an der Wunde ankommen gehören Antikörperzellen wie Makrophagen, welche die Wunde reinigen sollen. Durch die Erhöhung des Flüssigkeitsdurchsatzes, bildet sich eine Wundsichelung, auch Ödem genannt. Verschiedene weitere Körperstoffe sorgen für den Abbau von abgestorbenen Zellen in der Wunde und sorgen für viel austretendes Material, auch Exsudat genannt. Zudem wird in der Wunde eine Entzündung hervorgerufen, um weitere Bakterien und Viren abzuwehren. Durch diesen, hier stark vergrößerten, Ablauf, soll eine Wunde möglichst gut gereinigt werden und vor einer Entzündung durch Bakterien und Viren schützen. [16] [19]

Im Anschluss an die exsudative Phase beginnt die Proliferationsphase, auch Granulationsphase genannt. Diese setzt ungefähr ab dem vierten Tag nach dem Zuziehen der Wunde an. Das zuvor stark vertretene Fibrin wird in Verbindung mit anderen Leitstoffen genutzt, um die regenerativen Stoffe in die Wundfläche hineinzutragen. Dadurch kann die Regeneration von Gewebe beginnen. Dafür werden neue Gefäße ausgebildet und bereits existierende Gefäße gespalten, wobei die Geschwindigkeit dieser Vorgänge von vielen Faktoren wie zum Beispiel Alter, Gesundheit, Genetik abhängen. Zudem können externe Effekte wie nekrotisches Gewebe, Bakterien, Viren oder ein falsches Klima negative Effekte auf die Wundheilung in diesem Stadium haben. Dadurch kann sich die zweite Phase deutlich verlängern. [16] [19]

In der epithelialen Regenerationsphase bildet sich eine Matrix auf der oberen Schicht des Granulationsgewebes aus der vorherigen Phase. Sogenannte epitheliale Zellen verschließen die Wunde. Zudem reduziert sich in dieser Phase der Flüssigkeitsanteil der Wunde deutlich, um später Narbengewebe bilden zu können. [16] [17] [19] [20]

In der letzten Phase, der Remodellierungsphase, wird das in der vorherigen Phase erstellte Bindegewebe mit Narbengewebe überzogen. Dabei wird die Präsenz von Stoffen wie Kollagen an die Menge von gesunder Haut angepasst um neue, möglichst äquivalente Haut im Vergleich zur vorherigen Haut zu erhalten. Dieses Gewebe unterscheidet sich zur normalen Haut durch fehlende Elastizität. Bei geringer Tiefe der Wunde und gutem Heilungsverlauf bleibt keine sichtbare Narbe zurück. [16] [17] [19]

Der erfolgreiche Ablauf dieses Heilungsprozesses hängt von vielen verschiedenen Faktoren ab. Besonders bei älteren und vorerkrankten Patienten kann sich die Wundheilung von wenigen Wochen auf mehreren Monate erweitern. Diese dann als chronisch eingestuften Wunden bedürfen besonderer Behandlung und Pflege. [1] [18]

### 3.1.3 Chronische Wunden

Eine Wunde wird als chronisch eingestuft, wenn der Behandlungszeitraum acht Wochen überschreitet [1]. Dieser Fall tritt ein, wenn die Wunde an mindestens einer der vier Heilungsphasen ins Stocken gerät, sich entzündet oder zurückbildet [18] [21]. Die Gründe für chronische Wunden sind vielfältig, da der zugrundeliegende Heilungsprozess bereits äußerst komplex ist und das Zusammenspiel vieler körpereigene Stoffe an den richtigen Stellen der Wunde benötigt wird. Daher wird nachfolgend eine Auswahl von wichtigen, negativen Einflüssen auf die Wundheilung vorgestellt. Um die körpereigenen Stoffe nutzen zu können, muss nekrotisches und Schorfgebebe von der Wunde entfernt werden [17]. Diese blockieren die Wundfläche und verzögern den Heilungsprozess. Diese Blockade kann zu einer Unterversorgung der Wunde mit wichtigen Mineralstoffen und Sauerstoff führen. Zudem kann sich durch das zusätzliche Gewebe auf der Wunde Wasser ansammeln. Dies führt zu mehreren weiteren Effekten. Zum einen sorgt eine geringere Sauerstoffversorgung zu mehr nekrotischem Gewebe. Des Weiteren kann der Körper Stoffe des Immunsystems nicht richtig einsetzen, da diese Sauerstoff zur Aktivierung auf der Wunde benötigen [21]. Wenn die Wunde nicht von allen Bakterien befreit ist, kann diese sich entzünden und weitere Verzögerungen und Komplikationen hervorrufen [16]. Zudem gibt es weitere, systematische Ursachen, die eine schlechtere Wundheilung verursachen. Zum einen sorgt eine schlechte Ernährung, die zu Nährstoffmängeln führt, für negative Effekte auf die Heilung [16] [21]. Zum anderen können zunehmendes Alter, andere Krankheiten, Medikamente oder Stoffwechselstörungen zu einer verminderten Heilungsgeschwindigkeit führen [16].

Viele der Wunden, die sich zu chronischen Wunden entwickeln, fallen in drei Kategorien, welche daher der Fokus dieser Arbeit darstellen. Im nachfolgenden Abschnitt werden die Wundtypen erklärt und mit Bildern zur besseren Visualisierung ergänzt. Eine weit verbreitete Wunde ist der Dekubitus. Dieser Wundtyp entsteht durch längere Druckeinwirkung auf das Gewebe und tritt überall auf, aber besonders oft an Stellen des Körpers bei denen der Knochen nahe an der Haut liegt. Diese Körperstellen beinhalten zum Beispiel Fersen, Knöchel oder Ellbogen. Begünstigende Einflüsse für einen Dekubitus sind neben der Druckeinwirkung Scher- und Reibungseinwirkungen. Wird der Druck nicht schnell entfernt führt die vorherrschende Ischämie, so beginnen die Gewebezellen abzusterben und nekrotisches Gewebe entsteht. Anhand der erkennbaren Tiefe der Wunde lassen sich vier Schweregrade der Wunde festhalten, welche in Tabelle 3.1 erklärt werden. Der Dekubitus hat abhängig vom Schweregrad eine unterschiedliche Farbe. Von Rot über Pink bis Gelb und Schwarz, wenn Schorf oder eine Nekrose vorliegt. Der Rand der Wunde kann durch die Druckeinwirkung gerollt aussehen und darunter Tunnel oder Untermienungen vorhanden sein. Besonders ab dem dritten Schweregrad treten diese Erscheinungen auf, da die Wunde dann eine deutliche

### 3 Rigor

---

Tiefe hat, welche die Grundlage bietet. [15] [16] [18]

Tabelle 3.1: Tabelle mit Erklärungen für die vier üblichen Einordnungen von Dekubitus.

Grad	Erklärung
1	Nicht wegdrückbare Rötung der ansonsten gesunden Haut.
2	Partieller Verlust der oberen Gewebebeschicht. Oft rot-pinkes Wundbett.
3	Kompletter Verlust des Hautgewebes bis zum Muskelgewebe.
4	Kompletter Verlust des Hautgewebes bis zu den Knochen, Sehnen oder Organen.



Abbildung 3.1: Dekubitus mit einer partiellen Nekrose [22].

Die zweite in dieser Arbeit behandelte Wunde ist das diabetische Fußsyndrom, welches bei Diabetespatienten auftreten kann. Die Wunde muss unterhalb des Knöchels auftreten um in diese Kategorie zu fallen. Die auch als diabetischer Fuß bezeichnete Wunde tritt in drei verschiedenen Arten zutage, welche teilweise unterschiedliche Ursachen haben können (vgl. Tabelle 3.2). Beim neuropathischen diabetischen Fußsyndrom entstehen Wunden ähnlich wie beim Dekubitus aufgrund von Druckeinwirkungen oft auf der Sole des Patienten. Durch das fehlende Gefühl im Fuß, welches durch eine sogenannte Neuropathie, die durch Diabetes hervorgerufen wird, entsteht, werden diese Wunden zum Teil nicht wahrgenommen und entzünden sich daher häufig. Dies führt zu einer deutlichen Verschlechterung der Wundsituation beim Patienten. Die Ulzera, die beim Patienten auftreten, können auch Untermienungen und Tunnel aufweisen. Während der neuropathische diabetische Fuß sich durch ein fehlendes Gefühl im Fuß auszeichnet, ist der ischämische Fuß sehr schmerhaft. Wunden entstehen meist an den Zehen des Fußes, welches sich auch entzünden können. Das neuropathische ischämische diabetische Fußsyndrom wird durch eine rötliche Färbung des Fußes sichtbar

und erzeugt beim Tragen von engen Schuhe offene Ulzera mit Schorf. Dieser Wundtyp kann zudem zu Nekrosen im Fuß und besonders an den Zehen führen. [16] [18] [23] [24]

Tabelle 3.2: Tabelle mit Einordnung der drei Arten von diabetischen Fußsyndrom.

Art	Erklärung
Neuropathisch	Verlust des Gefühls in den Füßen durch Schädigung der peripheren Nerven. Oft mit schwer heilenden Ulzera verbunden.
Ischämisch	Durchblutungsstörung des Fußes welche zu einem stark schmerzenden Fuß führen kann.
Neuropathisch ischämisch	Zur Schädigung der peripheren Nerven kommt eine Durchblutungsstörung hinzu. Oft Nekrose oder Infektion von Ulcera verbunden.



Abbildung 3.2: Ulkus infolge einer Erkrankung an diabetischem Fußsyndrom [22].

Die letzte Kategorie von Wunden, die in dieser Arbeit behandelt wird, stellen die venösen und arteriellen Ulzera dar. Diese beiden Wundtypen werden in dieser Arbeit in eine Kategorie zusammengefasst, da die beiden zugrundeliegenden Krankheiten oft zusammen auftreten [25]. Damit ein venöser Ulkus auftritt muss eine chronische Veneninsuffizienz als zugrundeliegende Krankheit beim Patienten vorliegen. Für diese Krankheit gibt es viele verschiedene Gründe wie voranschreitendes Alter, Diabetes, fehlende Mobilität oder auch Venenthrombose. Venöse Ulzera finden sich meist an Stellen wie den Fußknöcheln und den Waden. Die Ulzera entstehen, wenn die Krankheit für längere Zeit besteht und zu Schäden an den Venen führt, welche dann zu Gewebeschäden führen (vgl. Tabelle 3.3). Bei arteriellen Ulzera liegt der Wunde eine arterielle Insuffizienz zugrunde. Diese führt zu einer Unterversorgung des Gewebes mit Nährstoffen und Sauerstoff und damit häufig zu einer starken

Nekrose. Mögliche Ursachen für diese Krankheit sind arterielle Thrombose oder arterielle Embolie. Aus der Krankheit heraus bilden sich arterielle Ulzera meist an den Zehen und der oberen Seite des Fußes bis zu Knöcheln, welche sich zudem häufig infizieren. Jedoch treten die Ulzera erst in der vierten Klassifizierung der Krankheit auf (vgl. Tabelle 3.4). Die venösen Ulzera können verschiedenen Größen annehmen, bleiben dennoch eher flach und sind oft mit Fibrin und Schorf bedeckt. Zudem sammelt sich um die Wunde Wasser an und das Gewebe färbt sich dunkel bräunlich (vgl. Abbildung 3.3). Das angesammelte Wasser führt auch zu einem starken Austreten von Flüssigkeit aus der Wunde. Arterielle Ulzera führen meist zu einer starken Nekrose, welche zudem durch dunkle Verfärbung des umliegenden Gewebes begleitet wird, und möglichen Wundbrand. Durch die starke Nekrose reicht die Wunde meist bis zum Muskelgewebe oder sogar Knochen. Die Wunde ist eher trocken und wirft kaum Flüssigkeit aus. Zudem verliert die Haut die Behaarung und wirkt glänzend (vgl. Abbildung 3.4). [15] [16] [18] [25] [26] [27] [28] [29]

Tabelle 3.3: Tabelle mit Erklärungen für die sieben üblichen Einordnungen von venösen Ulzera anhand der CEAP Klassifizierung [28] [27].

<b>Grad</b>	<b>Erklärung</b>
0	Keine äußereren Anzeichen von Veneninsuffizienz.
1	Teleangiektasie (sichtbare Kapillargefäße), wie Spinnenvenen .
2	Krampfadern
3	Flüssigkeitsansammlung am Ulkus
4	Hautgewebeschäden durch Venenstörung
5	Geheilten Ulzera
6	Aktiver venöser Ulkus

Tabelle 3.4: Tabelle mit Erklärungen für die vier üblichen Einordnungen von arterieller Insuffizienz [16] [26]

<b>Grad</b>	<b>Erklärung</b>
1	Symptomfrei.
2	Schmerzen in der Gliedmaße a) nach unter 200m Bewegung b) nach 200m Bewegung
3	Schmerzen in der Gliedmaße bei Ruhe
4	Ulkus mit Nekrosen oder Wundbrand an der Gliedmaße



Copyright Medetec (<http://www.medetec.co.uk>)

Abbildung 3.3: Ulkus infolge einer Veneninsuffizienz [22].



Abbildung 3.4: Leichter Ulkus aufgrund einer arteriellen Insuffizienz [29].

## 3.2 Maschinelles Lernen

### 3.2.1 Grundlegende Begriffe

#### Maschinelles Lernen

ML beschreibt den Prozess algorithmischen Lernens, durch welchen aus einem Datensatz Informationsmuster extrahiert werden, um eine Verbindung zwischen Eingabe und Ausgabe zu berechnen. Diese Prozess führt dann zu einer Generalisierung des Algorithmus, welcher sich vom algorithmischen Memorieren dadurch unterscheidet, dass nicht nur für gelernten Eingabedaten Ergebnisse berechnet werden können, sondern auch für neue und unbekannte Datenpunkte [30]. Wie in allen Bereichen der Künstlichen Intelligent, spaltet sich ML anhand des Ergebnisraumes in zwei Kategorien von Problemen. Dieser kann kontinuierlich oder diskret sein. Bei einem kontinuierlichen Ergebnisraum spricht man von einem Regressionsproblem, sonst von einem Klassifizierungsproblem. In Abbildung 3.5 ist ein Regressionsproblem dargestellt. Der originale Datensatz, auch Trainingsdaten genannt, ist durch die blauen Punkte dargestellt. Nun muss das ML Modell eine Funktion finden, welche alle bekannten Punkte möglichst gut annähert und gleichzeitig eine möglichst passende Darstellung der dazwischenliegenden Wertebereiche ermöglicht. Um dies zu erreichen, müssen zuvor Annahmen getroffen werden, wie die Lösungen für das Problem aussehen dürfen [31]. Im Bereich des ML wird meist angenommen, dass diese Funktion ein Polynom ist [30]. Anhand dieses Polynoms kann man dann alle nicht vorhandenen Werte berechnen. Abbildung 3.6 zeigt ein mögliches Klassifizierungsproblem im Bereich ML. Das System hat drei Ausgabewerte, welche den drei farbigen Gruppen von Datenpunkten entsprechen. Anhand des t Werts sollen dann neue Datenpunkte einer der Klassen zugeordnet werden. Je nach angewandtem Algorithmus sieht die Zuordnung anders aus und auch Datenpunkte, die keiner Klasse zuordenbar sind müssen berücksichtigt werden. In der bildbasierten Objektklassifizierung wird neben der Einordnung eines Objekts in eine der vorgegebenen Klassen, zudem noch eine Lokalisierung des Objekts gefordert. Dafür werden Begrenzungsboxen (BB) benutzt, welche die Fläche des Objekts im Bild umrunden. Dies können, wie in dieser Arbeit, einfache Rechtecke sein, aber auch komplexerer, unregelmäßige Polygone. Jede Begrenzungsbox soll dabei das Objekt möglichst genau umranden und nur wenig Zusatzfläche verwenden. Je komplexer das Polygon, desto weniger Zusatzfläche wird benötigt, aber umso aufwendiger ist die korrekte Annotation von Trainingsdaten, welche im nachfolgenden Abschnitt im Rahmen des überwachten ML beschrieben wird.

#### Arten des Lernens

Beide vorangegangenen Beispiele stellen sogenanntes überwachtes ML (engl. supervised ML) dar. Bei diesem Verfahren werden dem Algorithmus bereits korrekt Ergebnisse für die Trainingsdaten bereitgestellt. Aufgrund dieser Grundwahrheit versucht der Algorithmus eine Funktion zu finden, welche diese Eingabe- und Ausgabedaten abbildet. Im Kontrast zu überwachtem ML, lernt unüberwachtes ML (engl. unsupervised ML) eigene Muster aus den Trainingsdaten [32]. Diese erkannten Muster hängen vom angewandten Algorithmus und

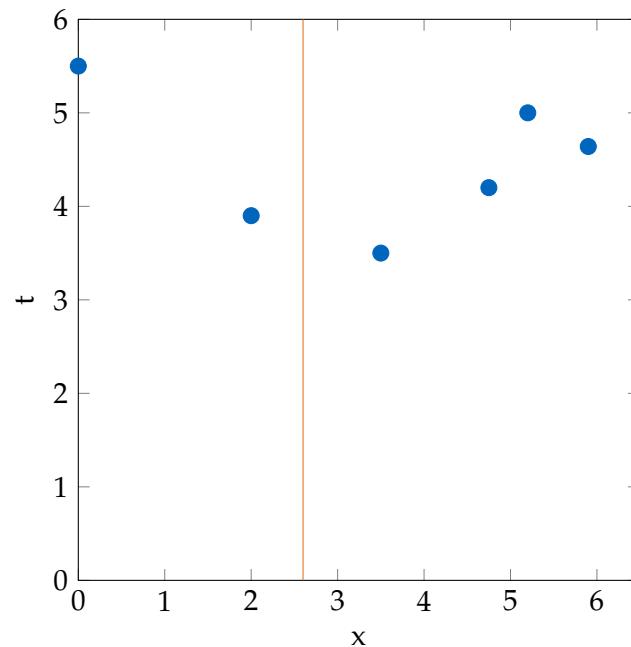


Abbildung 3.5: Beispielhafter Graph für ein überwachtes ML System mit Regressionsproblem.  
Ziel ist es  $t$  für den orangen  $x$  Wert zu erhalten, sodass es auf einem möglichst generalisierenden Polynom für die blauen Trainingsdaten liegt.

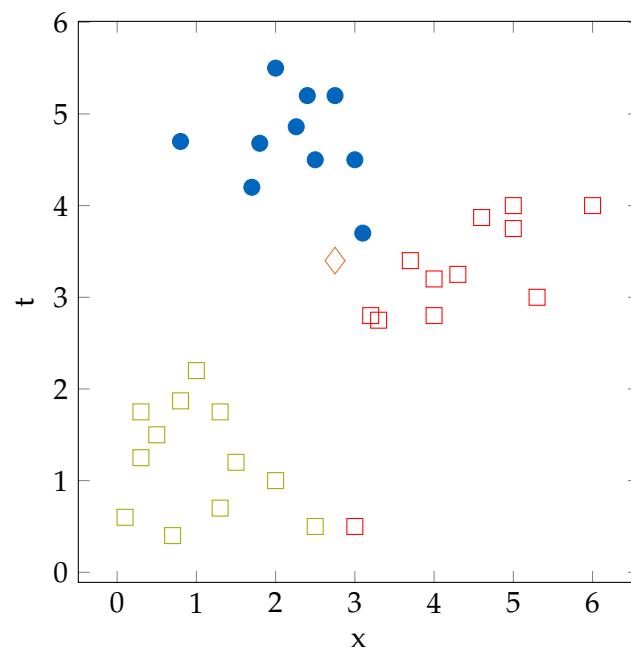


Abbildung 3.6: Beispielhafter Graph für ein ternär-klassifizierendes überwachtes ML System.  
Ziel ist es das orange  $x$  mithilfe der grünen, roten und blauen Trainingsdaten korrekt zu klassifizieren.

---

den benutzen Daten ab. Dabei gibt der Entwickler die Kontrolle über die gelernten Muster an den Algorithmus ab und muss anschließend überprüfen, ob das Modell die gewünschten Ergebnisse liefert. Eine der wichtigsten Methoden für diese Art von ML ist das Clustering von Daten, die im Werteraum nahe beieinander liegen. Die beiden vorangegangenen Lernmethoden sind auf den vorhandenen Datensatz beschränkt und können nie besser als der Datensatz werden, da nicht mehr Informationen vorhanden sind. Um diese Beschränkung zu überwinden wird bestärkendes Lernen (engl. reinforcement Learning) verwendet. Dabei führt das ML Modell seine Aufgabe wiederholt aus und erhält je nach Ergebnis eine positive oder negative Entlohnung. Dies führt dazu, dass das Modell deutlich mehr Situationen mit entsprechend korrekten Entscheidungen lernen kann, als an Datenpunkten vom Entwickler bereitgestellt werden können. Ein oft genutztes Beispiel sind klassische Brettspiele, wie Schach oder Go, bei denen das System durch einen Sieg ein positives und ansonsten ein negatives Feedback erhält [33]. Diese Spiele haben zu viele Konfigurationen um manuell jeden Zug bewerten zu können. Daher wird diese Methode des Lernens verwendet um möglichst viele Spielsituationen und anschließende Züge zu lernen.

Im Bereich des maschinellem Lernen existiert eine vierte Form des Trainings die auf den zuvor beschriebenen Formen aufsetzt. Das Transferlernen beschreibt die Spezialisierung eines breit gefächerten Modells auf einige wenige Aufgaben. Solche generalisierten Modelle können mit Bilddatenbanken wie Open Images [34] oder COCO [35] erstellt werden. Open Images, welches erstmals 2016 vorgestellt wurde, beinhaltet zum aktuellen Zeitpunkt 600 Klassen von Objekten. Der COCO Datensatz enthält zum aktuellen Zeitpunkt 330000 Bilder und 170 Klassen an Objekten. Wenn ein ML Modell auf einem solchen Datensatz trainiert wird, dann kann es viele verschiedene Muster und Objekte erkennen. Dies wird dazu genutzt, um die Klassifizierung der eigentlichen Objektklassen, die nicht bereits im Modell vorhanden sein müssen, zuverlässiger zu trainieren. Dabei können Teile des Modells fixiert werden, damit diese sich beim weiteren Training nicht mehr verändern. Zudem fügt man neue Elemente zu dem existierenden Modell hinzu, damit die Charakteristiken der neuen Klassen von Objekten gelernt werden [36]. Transferlernen wird besonders für Aufgaben genutzt die eine dünne Datenlage haben, da normalerweise tausende von annotierten Bildern benötigt werden um ein neues ML Modell zu trainieren. Zudem wirkt sich die große Datenbasis positive auf das Problem der Überanpassung an, welche in Abschnitt 3.2.2 genauer beschrieben wird.

## Evaluation von Modellen

Nachdem alle Trainingsschritte abgeschlossen sind, werden Messwerte zur Performanz des Modells benötigt. Daher gibt es eine Menge von Messwerten, die für verschiedene Situationen herangezogen werden können und so einen Vergleich mit anderen Modellen ermöglichen. Dafür müssen alle Testbilder zunächst manuell annotiert werden um als Grundwahrheit für die Evaluation dienen zu können. Die Messwerte, sowie deren Anwendungsgebiete werden nachfolgend erläutert. Alle Messwerte legen dieselben vier Einordnungen von Vorhersagen eines ML Modells zugrunde. Diese sind in Abbildung 3.7 farblich dargestellt. Die beiden grünen Kategorien 'Richtig Positiv' (RP) und 'Richtig Negativ' (RN) sollen maximiert werden,

da dort alle Vorhersagen hineinfallen, die mit der Grundwahrheit der Bilder übereinstimmen. Die rot markierten Felder stellen falsche Einordnungen dar. Vorhersagen die in die Kategorie 'Falsch Negativ' (FN) fallen, enthalten nicht die in der Grundwahrheit vorhandene Klassifizierung. Bei 'Falsch Positiv' (FP) vorhergesagten Bildern wurde eine Wunde klassifiziert, die nicht der Grundwahrheit vorkommt. Ein Problem mit dieser Klassifizierung stellt die

		Grundwahrheit	
		Wunde	keine Wunde
Vorhersage	Wunde	Richtig Positiv	Falsch Positiv
	keine Wunde	Falsch Negativ	Richtig Negativ

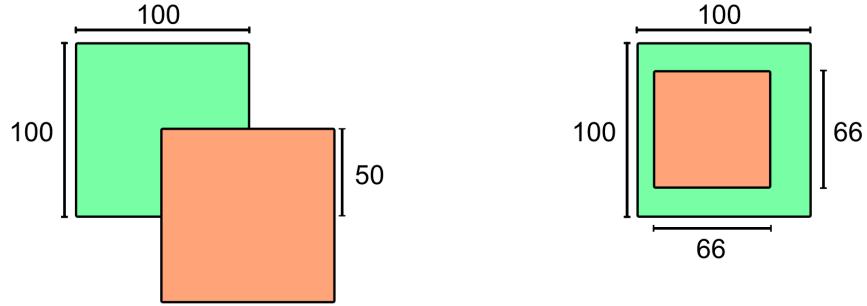
Abbildung 3.7: Wahrheitsmatrix mit den möglichen Einordnungen von Klassenvorhersagen zur Grundwahrheit eines ML Modells.

Einordnung einer Vorhersage in die vier Klassen dar. In dieser Arbeit fallen alle Vorhersagen, die keine Wunde oder einen falschen Wundtyp beinhalten, in die Zeile 'keine Wunde' und alle Vorhersagen, in denen eine Wunde mit dem korrekten Typ erkannt wird, in die Zeile 'Wunde'. Dadurch lassen sich gesamtheitliche Aussagen über das Modell treffen. Die meisten Objektklassifizierungsmodelle ordnen einer Klassifizierung zusätzlich einen Konfidenzwert zu, welcher angibt zu wie viel Prozent sich das Modell sicher ist, dass das erkannte Objekt von einem bestimmten Typ ist. Dieser Wert wird zusätzlich als Schwellwert herangezogen, sodass das Modell mindestens so sicher sein muss, dass das Objekt vom erkannten Typ ist. Damit eine Wunde als erkannt gilt, muss, wie zuvor erwähnt, der Typ mit einer Mindestkonfidenz übereinstimmen, zudem aber auch die erkannte Fläche der Wunde. Denn eine korrekter Wundtyp der an einer anderen Stelle des Bildes erkannt wird, als es in der Grundwahrheit festgehalten ist, kann nicht als korrekt anerkannt werden. Um solche fehlerhaften Klassifizierungen zu vermeiden, müssen die Flächen der erkannten Begrenzungsbox und der echten Begrenzungsbox miteinander verglichen werden. Eine oft verwendete Metrik stellt die Überschneidung über Vereinigung (IoU), auch Jaccard Index genannt, dar [37]. Dieser Wert zeigt an, wie viel Fläche der wahren und vorhergesagten Begrenzungsbox im Verhältnis zur Gesamtfläche der Beiden überlappen.

$$IoU = \frac{\text{Fläche Grundwahrheit} \cap \text{Fläche Vorhersage}}{\text{Fläche Grundwahrheit} \cup \text{Fläche Vorhersage}} \quad (3.1)$$

Die Formel ist in Gleichung 3.1 dargestellt und anhand von zwei Beispielen in Abbildung 3.8

visualisiert. Ein großer Vorteil dieser Metrik ist die Unabhängigkeit von der Skalierung der beiden Begrenzungsboxen. Wenn Beide mit dem gleichen Faktor skaliert werden, so verändert sich der IoU Wert nicht. Mithilfe der IoU und dem vorhergesagten Wundtypen lassen sich



$$IoU = \frac{50 \times 50}{2 \times 100 \times 100 - 50 \times 50} \approx 0.143$$

$$IoU = \frac{66 \times 66}{100 \times 100} = 0.4356$$

Abbildung 3.8: Zwei Beispiele für die Berechnung der IoU für die BB der Grundwahrheit (Grün) und der vorhergesagten BB (Orange).

alle Bilder in die Wahrheitsmatrix einordnen und erlauben die Berechnung der eigentlichen Performanzmesswerte. Ein oft angegebener Messwert ist die Genauigkeit eines Modells. Diese berechnet sich wie in Gleichung 3.2 und zeigt an wie viele der Vorhersagen des Modells im Verhältnis zu allen Vorhersagen korrekt sind.

$$\text{Genauigkeit} = \frac{RP + RN}{RP + RN + FP + FN} \quad (3.2)$$

Damit dieser Messwert eine gute Aussagekraft hat, sollten alle Klassen von Objekten in ihrer Anzahl an Datenpunkten ausgeglichen sein. Wenn eine starke Ungleichheit existiert, erreicht das Modell eine höhere Genauigkeit indem es immer die am stärksten vertretene Klasse von Objekten vorhersagt. Daher wird dieser Wert im medizinischen Bereich selten verwendet, da Krankheiten oft nur eine kleine Prozentzahl von Patienten betrifft und damit die Vorhersage 'ist nicht krank' meist über 90% Genauigkeit mit sich bringt. [38]

Um Aufgaben mit nicht balancierten Datensätzen bewerten zu können, werden andere Messwerte benötigt. Einfach zu evaluierende Messwerte sind die Präzision und der Recall.

$$\text{Präzision} = \frac{RP}{RP + FP} \quad (3.3)$$

$$\text{Recall} = \frac{RP}{RP + FN} \quad (3.4)$$

$$F1 = 2 \times \frac{\text{Präzision} \times \text{Recall}}{\text{Präzision} + \text{Recall}} \quad (3.5)$$

Die Präzision misst welchen Anteil der positiven Vorhersagen in der Grundwahrheit auch positiv sind. Dem entgegen misst der Recall welcher Anteil der positiven Grundwahrheit als positiv vorhergesagt wurde. Es sollten immer beide Messwerte evaluiert werden, damit ein besserer Einblick in die Performanz sichtbar ist. Diese beiden Werte können zu einem harmonischen Mittel, dem F1-Wert, zusammengeführt werden. Dieser Wert bewertet Präzision und Recall als gleichwertig, wie in Gleichung 3.5 zu sehen. Besonders bei Aufgaben die keine 'Richtig Negativ' Evaluation benötigen bietet sich dieser Wert an, da diese Einordnung von Vorhersagen nicht beachtet wird. Neben dem F1-Wert ist es sinnvoll die Präzision und den Recall trotzdem anzugeben, damit für Außenstehende ersichtlich ist, wie der Gesamtwert zusammenkommt. Bei nicht binären ML Modellen kann der F1-Wert auf die einzelnen Klassen oder auf das gesamte System angewendet werden. Dies wirkt sich auf das Gesamtergebnis aus, da ein niedriger Wert in der F1 Formel das Ergebnis durch die Multiplikation im Zähler deutlich verschlechtert. [39]

Zwei weitere wichtige Messwerte für binäre Klassifizierungsmodelle sind die Receiver operating characteristic (ROC) Kurve und die Fläche unter der Kurve (AUC). Die ROC Kurve ist ein Graph der die richtig positiv Rate (RPR) gegen die falsch positiv Rate (FPR) für verschiedenen Konfidenzniveaus plottet. Die AUC berechnet sich dann aus der Fläche der unterhalb des Graphen, wie in Abbildung 3.9 dargestellt. Die RPR ist der zuvor beschriebene Recall und die FPR wird mit der der Formel in Gleichung 3.6 berechnet.

$$FPR = \frac{FP}{FP + RN} \quad (3.6)$$

Für die ROC Kurve ist ein Verlauf oberhalb der Mittelkurve gewünscht. Denn die Mittelkurve stellt die ROC Kurve eines Systems dar, das zufällig zwischen den beiden Möglichkeiten, zum Beispiel 'Wunde' und 'keine Wunde', auswählt. Die AUC stellt eine Zusammenfassung aller Werte der ROC Kurve für die verschiedenen Konfidenzniveaus dar. Je stärker sich die AUC an den Wert 1 annähert, desto besser ist das ML Modell. Die AUC wird oft verwendet, verschiedene Klassifizierungsmethode miteinander zu vergleichen, während die ROC Kurve dazu genutzt werden kann, ein Konfidenzniveau für das Modell auszuwählen. [38]

Die ROC Kurve weiß allerdings bei der Aufgabe der Klassifizierung und Lokalisierung eine große Schwäche auf. Denn die Einordnung von Objekten in die richtig negativ Kategorie ist für dieses Problem nicht klar definiert. Wenn eine Wunde in einem Bild erkannt wird, müssten alle anderen Objekte als richtig negativ eingetragen werden. Dafür müsste ein zweites Modell implementiert werden, dass alle anderen möglichen Objekte erkennen kann. Oft wird daher eine Spezifizierung der Kategorie richtig negativ vorgenommen. Da in dieser Arbeit solch eine Spezifizierung der Definition schwer anzubringen ist, wird statt der ROC Kurve eine Präzision-Recall Kurve verwendet, welche statt der RPR und FPR die Präzision

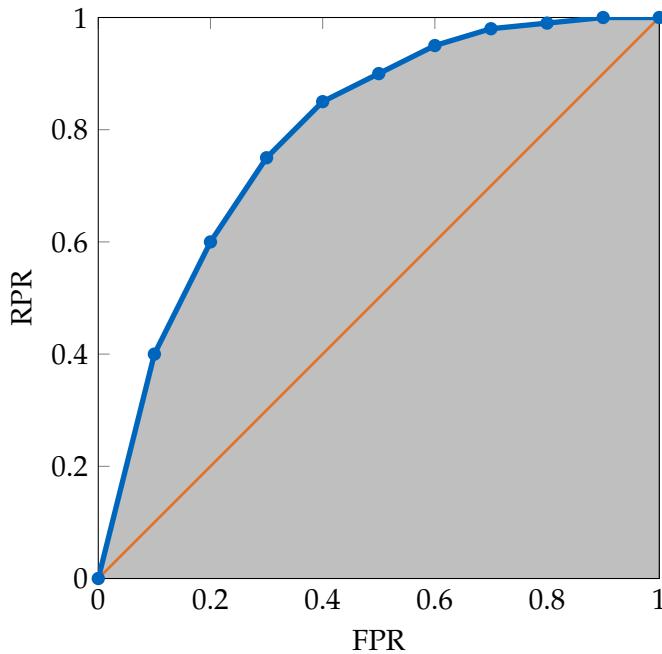


Abbildung 3.9: Beispielhafte ROC Kurve (blau), dazugehörige AUC (grau) und Mittelkurve (orange)

und den Recall als Achsen hat. Über diese Kurve ist ebenfalls eine AUC berechenbar, welche dann die viel gebrauchte gemittelte durchschnittliche Präzision (mAP) darstellt. Nachfolgend wird die AUC über die ROC Kurve in den vorgestellten Arbeiten genutzt und die AUC der Präzision-Recall Kurve speziell im Kontext von Kapitel 5 verwendet.

### 3.2.2 Grundlegende Algorithmen

#### Neuronale Netzwerke

In dieser Bachelorarbeit werden Convolutional Neural Networks (CNN) (deutsch: gefaltete neuronale Netzwerke) eingesetzt. Diese stellen eine Spezialisierung von neuronalen Netzwerken für Bilder dar. Im folgenden Abschnitt werden daher zuerst die Grundlagen von neuronalen Netzwerken und anschließend die Anpassungen für CNNs erklärt. Die Erklärungen basieren auf dem öffentlich zugänglichen Material der Vorlesung 'Convolutional Neural Networks for Visual Recognition' von Professorin Fei-Fei der Stanford Universität [40].

Das übergeordnete Ziel ist es, für zuvor deklarierte Klassen, Objekte auf Bildern erkennen zu können. Für diese Aufgabe kann ein Neuronales Netzwerk als Architektur für Maschinelles Lernen verwendet. Ein neuronales Netzwerk ist eine Verkettung von Funktionen, die in mehreren Ebenen organisiert sind (vgl. Abbildung 3.10). Es muss immer eine Eingabe-, Ausgabe- und mindestens eine versteckte Ebene geben. Die Eingabeebene nimmt das Bild entgegen. Die Ausgabeebene gibt die finale Zugehörigkeit zu den zu erkennenden Klas-

sen aus. Die versteckten Ebenen sind die lernenden Ebenen und werden trainiert um die gewünschten Objekte korrekt zu erkennen. Sie werden als versteckt bezeichnet, da durch die Überlagerung von vielen Funktionen einzelne Funktionsergebnisse nicht mehr aus dem Gesamtergebnis abgeleitet werden können. Jede Ebene besteht aus vielen sogenannten Neu-

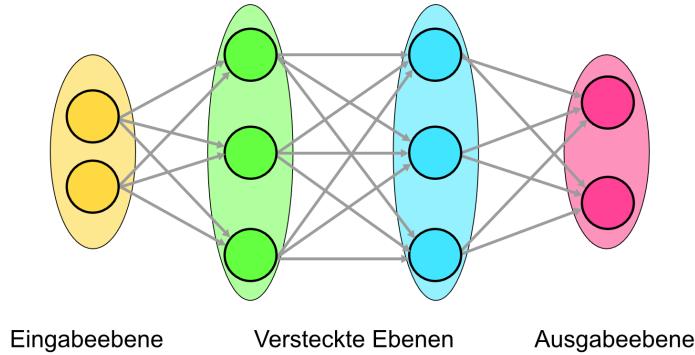


Abbildung 3.10: Darstellung eines 2-Ebenen Neuronalen Netzes mit vollständig verbundenen Ebenen. Enthält 2 verdeckten Ebenen mit drei Neuronen, eine Eingabeebene mit zwei Neuronen und eine Ausgabeebene mit zwei Neuronen

ronen, welche ein Eingabesignal empfangen, dann ein Skalarprodukt sowie eine Addition zwischen Eingabe und den eigenen Parametern berechnen und anschließend anhand einer Aktivierungsfunktion ein Ausgabesignal liefern. Die Neuronenparameter bestehen aus einer Matrix mit verschiedenen Werten, fortan Gewichte genannt, und einem Vektor von Werten welcher als Bias bezeichnet wird. Um ein Skalarprodukt zwischen der Eingabe und den Parametern durchführen zu können, benötigen sie die richtigen Dimensionen. Die Eingabe wird immer als eindimensionaler Vektor mit den Dimensionen [Datengröße, 1] dargestellt.

$$f(x, W) = \begin{pmatrix} 4 & 7 & 3 \\ 5 & 3 & 6 \end{pmatrix} \left| \begin{array}{c} 2 \\ 3 \\ 1 \end{array} \right. \cdot \begin{pmatrix} 5 \\ 2 \\ 3 \\ 1 \end{pmatrix} \quad (3.7)$$

Die Gewichtsmatrix hat die Dimensionen [Ausgabegröße, Datengröße]. Der Biasvektor hat die Dimensionen [Ausgabegröße, 1] und wird nach dem Skalarprodukt auf das Ergebnis aufaddiert. Häufig wird die Gewichtsmatrix und der Biasvektor zu einer Gewichts-Bias-Matrix zusammengeführt und der Eingabevektor in der Datengröße mit einer eins vergrößert, um die Addition in das Skalarprodukt zu integrieren. Dies ist in Gleichung 3.7 durch die zusätzlichen Trennstriche in der Matrix  $W$  und dem Eingabevektor  $x$  dargestellt. Die Gleichung zeigt eine beispielhafte Berechnung eines linearen Neurons in der Ausgabeebene. Da die Ausgabeebene die letzte Ebene darstellt, muss die Ausgabegröße der Anzahl zu lernenden

Klassen entsprechen. In diesem Beispiel handelt es sich um eine binäre Klassifizierung, da der Ergebnisvektor zwei Werte enthält, welche die Bewertung der Zugehörigkeit zu diesen Klassen darstellt. Im Gegensatz zur Ausgabeebene, wird in den versteckten Ebenen die Größe des Ausgabevektors anhand von Hyperparametern, sowie möglichen Farbkanälen von Bildern bestimmt. Hyperparameter sind Parameter die manuell festgelegt werden müssen, allerdings mit verschiedenen Testmethoden optimiert werden können. Die in Gleichung 3.7 dargestellte lineare Funktion stellt meist die grundlegende Berechnung des Neurons dar und wird durch nicht-lineare Aktivierungsfunktionen erweitert. Die Aktivierungsfunktion eines Neurons nimmt die Ausgabe des Skalarprodukts und berechnet eine Ausgabe, die dafür verwendet wird zu bestimmen mit welcher Rate die Daten an die nächste Ebene weitergeleitet werden. Eine stark vertretenen Aktivierungsfunktion stellt dabei die in Abbildung 3.11 gezeigte Rectified Linear Unit Aktivierungsfunktion dar. Diese gibt bei Eingabewerten kleiner oder gleich Null eine Null aus und ansonsten die Eingabe. Diese einfach zu berechnende Funktion ist besonders effizient das sie nur eine Operation benötigt. Allerdings haben verschiedene Aktivierungsfunktionen auch verschiedene negative Effekte. So führt das Ausgeben von Null für negative Eingabe zum möglichen 'Sterben' des neuronalen Netzwerkes, wenn keine positive Werte eingegeben werden. Denn dann werden keine Informationen mehr an die nächste Ebene weitergeleitet. Um die Neuronen zu bestimmen welche die Ausgabe

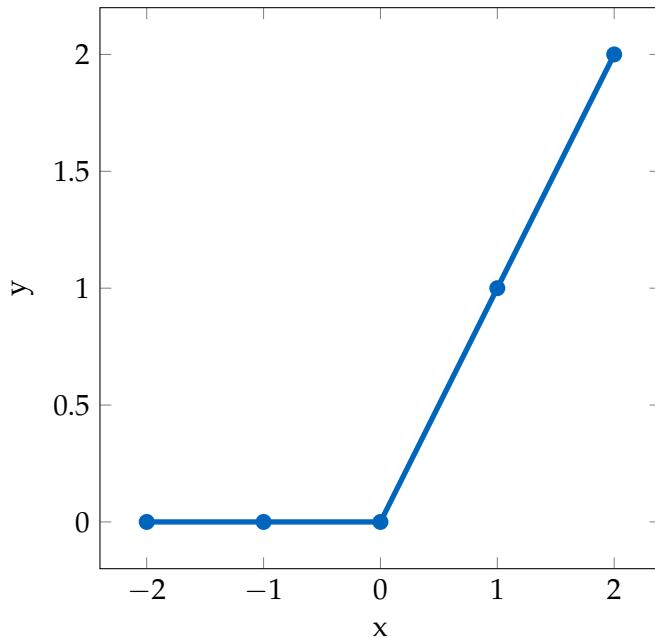


Abbildung 3.11: Rectified Linear Unit Aktivierungsfunktion

der Aktivierungsfunktion erhalten, werden alle Neuronen in Ebenen organisiert. Die drei Einteilungen in Eingabe-, versteckte und Ausgabeebenen wurde bereits zuvor erklärt und sind in Abbildung 3.10 zu sehen. Ebenen werden anhand ihrer Funktionalität und Verbindungen zudem einer Klasse zugeordnet. In Neuronalen Netzwerken werden besonders vollständig

verbundene Ebenen verwendet. Dies sind Ebenen die ihr Ergebnis auf den zu lernenden Parametern, den geltenden Hyperparametern und ihren Eingaben berechnen. Zudem sind alle Neuronen dieser Ebene mit allen Neuronen der nächsten Ebene verbunden. In vielen Fällen werden in Neuronalen Netzen nicht mehr als Drei oder Vier versteckte Ebenen verwendet. Der Grund für die geringe Zahl an Ebenen ist das Phänomen des Überanpassung.

Überanpassung beschreibt die Modellierung von Rauschen in einem Datensatz durch eine zu hohe Fähigkeit komplexe Funktionen zu approximieren. Mehr Ebenen und Neuronen in einem Netzwerk erlauben es komplexere Funktionen zu approximieren. Allerdings ist das Ziel eines neuronalen Netzes die Generalisierung. Wenn ein Netz alle Trainingsdaten perfekt erlernt, also eine womöglich sehr komplexe Funktion, dann wurden auch Datenpunkte, sogenanntes Rauschen, die abseits der Hauptdatenpunkte liegen modelliert. Dies führt dazu, dass das Modell die Datenübergreifenden Merkmale der Datenpunkte außer Acht lässt und stattdessen eine schlecht generalisierende Funktion durch den Datenraum legt. Dies ist beispielhaft in Abbildung 3.12 dargestellt. Die grüne Funktion ist überangepasst, da sie jeden blauen und roten Datenpunkt, der im Bereich der anderen Farbe liegt, dennoch durch komplexe Windungen in den korrekten Farbraum zuordnet. Werden nun Datenpunkte die nah an der beispielhaft gut generalisierenden Funktion in das Programm gegeben, würde die grüne Funktion einige Fehler machen. Um dies zu verhindern gibt es verschiedene Techniken die im Detail in [41] und [42] beschrieben werden.

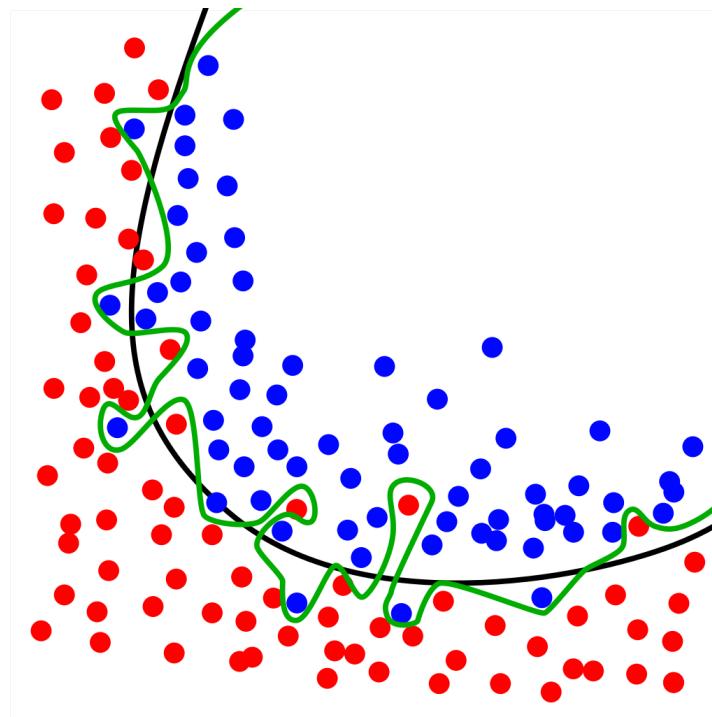


Abbildung 3.12: Darstellung einer überangepassten Funktion (grün) im Vergleich zu einer generalisierten Funktion (schwarz)[43].

## Convolutional Neural Networks

Convolutional Neural Networks stellen eine Spezialisierung von Neuronalen Netzen für Bilderkennung und Objektklassifizierung dar. Während Neuronale Netze mit wenigen Ebenen auskommen, gilt bei CNNs, dass mehr Ebenen zu besseren Ergebnissen führen. Auch hier muss auf Überanpassung mit verschiedenen Methoden geachtet werden, allerdings führen mehr Parameter und Ebenen nicht zu starker Überanpassung solange die Datengrundlage groß genug ist [44]. Ebenen in CNNs sind standardmäßig dreidimensional. Denn es werden oft farbige Bilder mit den drei Farbkanälen Rot, Grün und Blau verwendet. Im Gegensatz zu den ersten beiden Dimensionen, wird diese dritte Dimension durch das gesamte Netzwerk beibehalten. Die anderen beiden Dimensionen, Breite und Höhe, können mit neuen Ebenenklassen, die bei CNNs verwendet werden, verändert werden.

Die wichtigste Klasse von Ebenen sind die convolutional Ebenen. Diese besitzen sogenannte Filter, welche kleine, sich bewegende Fenster darstellen, die über das gesamte Bild bewegt werden. Diese Filter berechnen dabei in ihrem Einzugsbereich, zum Beispiel  $3 \times 3 \times 3$  Pixel, ein Kreuzprodukt mit ihren Parametern um eine zweidimensionale Aktivierungsabbildung des Bilds zu berechnen. Diese Abbildung zeigt dann an auf welche Art von Merkmalen der entsprechende Filter reagiert hat. Merkmale können Kanten, Farbmuster und andere Eigenschaften in Bildern sein. Kleinere Merkmale werden dann in späteren Ebenen zu größeren Merkmalen wie zum Beispiel einer Wunde zusammengesetzt. Jeder Filter wird von drei Hyperparametern beeinflusst, die angeben wie viele Filter pro Merkmal genutzt werden sollen, wie viele Pixel der Filter pro Schritt weitergeschoben wird und mit wie vielen Nullbits das Bild am Rand erweitert wird um die Eingabegröße konstant zu halten. Die Ausgabe der Ebene wird im Unterschied zu den vollständig verbundenen Ebenen nicht an alle Neuronen der nächsten Ebene weitergeleitet sondern nur eine lokale Anzahl von Neuronen die auch durch Hyperparameter bestimmt werden können.

Bei CNNs wird eine zweite neue Klasse von Ebenen verwendet. Diese werden Bündelungsebenen genannt. Dabei werden die Bilder in Höhe und Breite verkleinert um ausschließlich relevante Daten an die nächste Ebene zu senden und die Parameterzahl deutlich zu verringern und damit auch Überanpassung zu verhindern. In diesen Ebenen wird eine Funktion berechnet, die aus einem Filter nur einen gewissen Anteil an Informationen ausgibt. Eine weitverbreitete Funktion ist die Maximumsfunktion welche auf einen  $2 \times 2 \times 3$  Filter angewendet wird. Dadurch ist die Ausgabe 75% kleiner als die Eingabe. Die Bündelungsebenen bauen auf der Idee auf, dass es für die Merkmalerkennung reicht, eine grobe Position der Merkmals im originalen Bild zu haben statt die genauen Pixelkoordinaten. [45]

### 3.2.3 Verwendeter Algorithmus

Das in dieser Bachelorarbeit betrachtete CNN basiert auf einer Faster R-CNN Architektur. Diese wiederum basiert auf zwei CNNs, die zu einem CNN verschmolzen wurden und ein großes R-CNN zu bilden. Das als Detektor verwendete CNN basiert auf der Fast R-CNN Ar-

chitektur [46], welches eine Weiterentwicklung des originalen R-CNN aus dem Jahr 2014 von Girshick et al. [47] darstellt. Das zweite CNN ist ein Regionsvorschlag-Netzwerk (RPN), das speziell darauf trainiert wurde Regionsvorschläge für Objekte in Bildern zu berechnen [48]. Im folgenden Abschnitt wird die grundlegende Architektur des R-CNN von Girshick et al. aus dem Jahr 2014 beschrieben und anschließend die Veränderungen des Fast R-CNN und RPN, sowie deren Verschmelzung erklärt.

Die R-CNN Architektur setzt im Vergleich zu reinen CNN Architekturen auf vorhergehende Analyse der Bilder. Jedes Bild wird zunächst mithilfe des Selektive Suche Algorithmus von Uijlings et al. [49] auf mögliche Vorschläge für Regionen untersucht. Diese Regionen werden dann an das CNN weitergeleitet um pro Region mit einem CNN einen 4096-dimensionalen Merkmal-Vektor zu extrahieren. Fünf gefalteten und zwei vollständig verbundenen Ebenen führen diese Extraktion auf 227x227 Pixel mittel-subtrahierten RGB Bildern aus. Bei einem mittel-subtrahierten Bild wird für jeden Pixel der Mittelwert über alle Bilder an diesem Pixel subtrahiert. Bei einem RGB Bild wird dies separat für jeden der drei Farbkanäle durchgeführt. Das führt dazu, dass alle Bilder dieselbe Größe haben müssen. Im Anschluss werden die Bilder als Eingabe für Stützvektormaschinen (SVM) genutzt, welche jeweils für eine der zu erkennenden Klassen trainiert wurden. Daraus ergibt sich dann eine Bewertung der Zugehörigkeit zu den Klassen. Zum Schluss berechnet ein lineares Regressionsmodell für Begrenzungsboxen die finale Region, welche mit der Begrenzungsbox umrandet wird. Diese Neuberechnung der Region verbesserte im originalen Paper das finale Ergebnis. [47]

Das Fast R-CNN benötigt anders als das R-CNN ein Bild und die Regionsvorschläge als Eingabe um eine Detektion auszuführen. Die Regionsvorschläge werden als gegeben angenommen, und können zum Beispiel mit dem Selektive Suche Algorithmus berechnet werden. Um das gleiche Muster an verschiedenen Positionen des Bildes erkennen zu können, wird im ersten Schritt mithilfe von gefalteten und Maximum-bündelnden Ebenen eine Merkmal-Abbildung generiert. Diese Abbildung beschreibt eine Matrix, die auf alle Regionen eines Bildes angewendet wird, um zu prüfen ob das beschriebene Muster vorhanden ist [50] [47]. Im nächsten Schritt wird aus der Merkmal-Abbildung in den Regionsvorschlägen eine Merkmalvektor extrahiert und an mehrere vollständig verbundene Ebenen weitergeleitet. Die letzte Ebene leitet ihr Ergebnis an eine SoftMax-Ebene sowie eine Begrenzungsboxebene äquivalent zu R-CNN weiter. Die SoftMax-Ebene berechnet dabei mithilfe der SoftMax-Funktion die finalen Zugehörigkeiten zu den trainierten Klassen anhand von Wahrscheinlichkeiten. Die SoftMax-Funktion wandelt dabei einen K-dimensionalen Vektor mit beliebigen realen Werten in einen K-dimensionalen Vektor mit realen Werten zwischen Null und Eins um und garantiert dass die Summe über alle Vektorelemente Eins ergibt [51]. Das erlaubt es dem Detektor zu entscheiden welcher Klasse ein Objekt zugeordnet wird. Die Verbesserung des Fast R-CNN gegenüber dem klassischen R-CNN stellt dabei einerseits die einzelne SoftMax-Ebene gegenüber einer SVM pro Klasse von Objekten dar. Andererseits wird das CNN zur Berechnung einer Merkmal-Abbildung genau einmal für das gesamte Bild berechnet anstatt für alle Regionsvorschläge erneut. Dies spart deutlich Rechenzeit und verbessert die Detektionsergebnisse

leicht. Diese Ergebnisse gelten solange die Regionsvorschläge in identischer Geschwindigkeit und Genauigkeit berechnet werden wie bei R-CNN. [46]

Das Faster R-CNN Modell erweitert das zuvor beschriebene Modell mithilfe eines zweiten CNNs welches speziell für Regionsvorschläge gebaut ist. Dieses RPN erhält als Eingabe ein zuvor berechnete zweidimensionale Merkmalabbildung des Bildes und berechnet dann mithilfe eines bewegenden Fensters die einzelnen Objektregionen. Das klassische bewegte Fenster wird in RPN durch Ankerpunkte erweitert. Das erlaubt dem System leichter unterschiedliche Größen von Regionen auf Merkmale zu prüfen. Die Ankerpunkte befinden sich dabei im Zentrum des Fensters und werden genutzt um unterschiedliche Skalierungen von Begrenzungsboxen für die Regionsvorschläge zu testen. Das ermöglicht es leichter verschiedene Positionen und Rotationen im Bild als eine Region zu erkennen. Die entstehenden Regionsvorschläge werden im Anschluss mithilfe einer Bündelungsebene an das Fast R-CNN weitergeben. [48]

Eine weitere Besonderheit der Architektur kommt besonders beim Training des Systems zum Tragen. Denn die beiden CNNs werden nicht unabhängig voneinander trainiert sondern teilen sich versteckte Ebenen und trainieren diese gemeinsam. Dadurch werden beide CNNs auf gleicher Weise trainiert und unterscheiden sich nicht so stark auf der sonst zufälligen Initialisierung pro CNN. Der Prozess des Trainings beginnt bei dem RPN welches zuerst einen Trainingsdurchlauf macht. Im Anschluss werden die Ergebnisse für das Training des Fast R-CNN genutzt. Die Ergebnisse des Detektionsmodells werden dann wieder für das RPN genutzt. Dieser Zirkel wird während dem Training für einige Zeit durchlaufen. Damit das RPN keine Ebenen verändert, die es selbst nicht benötigt, werden diese Ebenen während dessen fixiert und können nicht geändert werden bis das Fast R-CNN diese trainiert. Zum Schluss des Trainings werden die Ebenen die exklusiv zum Fast R-CNN gehören für einige Zeit trainiert, um das Ergebnis des Detektors zu verbessern. Diese Architektur führt dazu, dass das Modell sehr gute Ergebnisse und eine hohe Trainingsgeschwindigkeit bietet. [48]

### 3.3 Weitere Einsatzmöglichkeiten von maschinellem Lernen in der Wundmedizin

Im vorangegangenen Abschnitt wurde die verwendete ML Technologie im Rahmen dieser Arbeit erläutert. Im Bereich der künstlichen Intelligenz und des ML gibt es in letzten Jahren einen deutlichen Zuwachs an Arbeiten im Bereich der medizinischen Bildverarbeitung. Im folgenden Abschnitt werden daher Forschungsarbeiten, die sich mit diesem Bereich beschäftigen vorgestellt.

Nach intensiver Suche und bestem Wissensstand zufolge gibt es nur eine Arbeit von Goyal et al. [52] welche sich mit der Klassifizierung von Wundtypen beschäftigt. In der Arbeit wird die Detektion von diabetischen Fußulkus mit fünf existierenden ML und CNN Architekturen sowie einem eigenen Modell untersucht. Dafür wurden 397 Bilder von gesunden und nicht

gesunden Füßen herangezogen und von medizinischem Personal annotiert. Die Annotationen wurden mit teils komplexen Begrenzungsboxen durchgeführt um jede Fläche der Wunden genau zu markieren. Zu den bereits existierenden Modelle zählen GoogLeNet [53], AlexNet [44] und LeNet [54]. Das neue Modell der Arbeit, fortan DFUNet genannt, ist im Gegensatz zu den anderen Modellen speziell mit dem Fokus auf diabetischen Fußulkus entwickelt worden. Dafür verwendet das CNN neben standardisierten Klassen von Ebenen auch parallele convolutional Ebenen, die mehrere Merkmalsextraktionen pro Region ermöglichen. Dafür werden drei unterschiedlich große Filter von 1x1, 3x3 und 5x5 Pixel verwendet, deren Ergebnis im Anschluss konkateniert und weitergeleitet wird. Die daraus resultierenden Ergebnisse sind bei der abschließenden Performanzmessung deutlich vor anderen Modellen. Das DFUNet erreicht bei der Sensitivität, der Genauigkeit, dem F-Wert, sowie dem AUC-Wert und Standardfehler die besten Ergebnisse. Bei der Spezifität und der Präzision liegt das Modell wenige Punkte hinter dem GoogLeNet Modell. Mit einem AUC-Wert von 0.961 und einem F1-Wert von 0.939 erreicht das Modell eine hohe Performanz.

Eine weitere Arbeit, die eine fachliche Überschneidung mit dieser Bachelorarbeit besitzt, ist die Schwereinstufung von diabetischen Wunden von Zhao et al. [55]. Das Team hat ein bilineares CNN entworfen welches anhand von Wundtiefe und Granulationsgewebe eine Bepunktung des Schweregrads der Wunde für den PWAT Fragebogen [56] erstellt. Ein bilineares CNN beschreibt dabei zwei Merkmalsextraktionspipelines, die jeweils auf einem CNN basieren. In dieser Arbeit wurde als Grundlage für beide Pipelines dasselbe CNN verwendet welches auf der VGG Architektur [57] basiert und mithilfe des ImageNet Datensatzes [58] vorab trainiert wurde. Dies erlaubt es Transferlernen einzusetzen und so die bereits gelernten Merkmale aus dem Datensatz auf die neuen Klassen von Bildern anzuwenden. Die Ergebnisse der beiden CNNs werden am Ende der Pipelines verrechnet und als bilinearer Vektor an vollständig verbundene Ebenen weitergeleitet. Das Modell erreicht im Bereich der Wundtiefe und auch bei der Menge an Granulationsgewebe eine Genauigkeit von 0.846.

Im Gegensatz zur Klassifizierung von Wundtypen lassen sich viele wissenschaftliche Arbeiten im Bereich der Wundflächenberechnung sowie Gewebeererkennung finden. Das Ziel der Arbeiten ist es ungenaue und mechanische Messungen an der Wunde des Patienten durch bildgestützte Flächenerkennung zu verbessern. Poon et al. [59] entwickelten eine Smartphone Applikation, die genutzt werden kann, um Bilder von einer Wunde zu machen und anschließend die Größe dieser Wunde detektiert sowie den Status der Wundheilung und Elemente der Wunde, wie zum Beispiel Nekrose, darstellt. Biswas et al. [60] haben eine SVM trainiert um pixelgenau verschiedene Wundgewebe zu extrahieren. In der Arbeit erreicht das Modell eine Genauigkeit von 0.7198 für Bilder, die in verschiedenen Lichtumgebungen aufgenommen werden können. Eine weitere CNN basierte Methode zur Segmentierung von Wunden wurde von Wang et al. [61] entwickelt. Das Modell soll als Framework dienen, um weitere Applikationen wie Infektionserkennung und Heilprozessvorhersage an die Wunderkennung anzuschließen. Mit einer Genauigkeit von 0.95 und einem durchschnittlichen IoU-Wert von 0.473 erreicht das Modell gute Klassifizierungen, auch wenn die Flächenüberlagerung zur Gesamtfläche mit

dem der Grundwahrheit weniger als 0.50 darstellt. Bei der Infektionserkennung erreicht das Modell eine Genauigkeit von 0.956 und einem AUC-Wert von 0.847. Weitere Messwerte sind nicht veröffentlicht worden.

Khalil et al. nutzen Gradient Boosted Trees um vier Gewebetypen von chronischen Wunden pixelgenau zu klassifizieren [62]. Um bessere Ergebnisse zu erzielen, werden in der Arbeit neben dem RGB-Farbraum, vier verschiedene andere Farträume, wie HSV oder normalisiertes RGB, untersucht. Zudem wurde versucht die aktuelle Stufe der Heilung einer Wunde anhand der klassifizierten Gewebeanteile zu bestimmen. Insgesamt erreicht das Team eine Genauigkeit von 96% mit Gradient Boosted Trees als ML Algorithmus.

Die vorgestellten Arbeiten stellen eine kleine Teilmenge der ML gestützten Arbeiten im Bereich der Wundpflege dar. Im nachfolgenden Abschnitt werden zudem nicht ML gestützte Arbeiten für die Wundpflege vorgestellt, um einen Einblick in andere Bereiche der Informatik zu geben und eine Grundlage für die Möglichkeit einer späteren Integration dieser Arbeit in andere Systeme zu bieten.

### 3.4 Weitere technische Hilfsmittel in der Wundpflege

Der Bereich der Wundpflege umfasst viele Gebiete der Informatik außerhalb der künstlichen Intelligenz. Viele der Arbeiten umfassen allgemeinere Systeme für die Wundpflege und bieten damit auch die Möglichkeit ML gestützte Systeme zu integrieren. Daher werden einige ausgewählte Arbeiten in diesem Abschnitt vorgestellt.

Ein viel untersuchtes Gebiet in der Wundpflege ist die Erstellung von Smartphone Applikationen, die durch den Patienten selbst genutzt werden können. Yee et al. stellen in ihrer Arbeit [63] eine App zur Organisation von Wundbilddaten sowie zur 3D Bildgenerierung von Wunden vor. Die 3D Bilder der Wunden bieten Genauigkeit bis in den Millimeterbereich und erlauben Ärzten eine einfache Inspektion von Wunden. Besonders mit Blick auf die voranschreitende Digitalisierung und Telemedizin ist solch eine Applikation von hohem Wert.

Eine weitere Applikation für Tablets und Smartphones wurde von Wang et al. erstellt [64]. Diese soll Krankenschwestern bei der Behandlung von Patienten helfen. Das Programm organisiert die Patientendaten in der Krankenhausdatenbank und gibt wichtige Hilfestellung im Prozessablauf. Nachdem die Krankenschwester sich mit dem bisherigen Verlauf sowie mögliche Notizen vorheriger Behandlungen genau über den Patienten informiert hat, kann sie mit dem Gerät Fotos machen, die Wunde darauf umranden damit das Programm danach selbstständig den Wundrand bestimmen und Metriken wie Fläche, Gewebeanteile und Fortschritt zu vorherigen Bildern berechnen kann. Anschließend können weitere Daten wie Behandlung und verwendete Verbände eingetragen werden. Die Applikation wurde mit 10 Krankenschwestern im laufenden Betrieb des Xiangya Krankenhauses in China getestet. Die Tests zeigen, dass die Krankenschwestern bei vielen Aufgaben schneller und genauer arbeiten

können als zuvor. So sind Messdaten wie die Wundfläche deutlich genauer als zuvor mit manuellen Messmethoden wie Linealen. Zudem sind alle Patientendaten zeitlich sortiert und mithilfe eines scannbaren Labels direkt aufrufbar. Das verhindert manuelles Suchen in der Patienten- und Bilddatenbank. Bei den anschließenden Befragungen der Krankenschwestern wurde hohe Zufriedenheitsniveaus gemessen, auch wenn die Krankenschwestern weitere Verbesserungsmöglichkeiten angemerkt haben.

Mit Blick auf die vorgestellten Arbeiten zeigt sich eine starke Fokussierung auf Geräte, die während der Behandlung in der Hand gehalten werden müssen. Dies bringt einige Probleme wie Bakterien, benötigte Ablageflächen und Wechsel zwischen Gerät und Patient mit sich. Daher hat Klinker et al. eine Applikation für die HoloLens, eine Mixed Reality Brille, entwickelt [65]. Eine solches Head-Mounted-Display wird auf den Kopf gesetzt und ermöglicht es virtuelle Anzeigen in das reale Blickfeld des Nutzer einzuspeisen. Zusätzlich sind Funktionen wie Positionstracking und Erkennung von größeren Objekten im Raum verfügbar. Die Applikation bietet die Möglichkeit eine Checkliste für die Wundbehandlung anzuzeigen, in welche neue Daten über die Wunde mit vorgefertigten Möglichkeiten eingetragen werden können, und die Wunde zu vermessen. In drei Design Cycles hat das Team verschiedene Versionen der Applikation mit jeweils neuen Verbesserungen implementiert. Ab dem zweiten Design Cycle wurde die HoloLens verwendet, zuvor noch das Vuzix Monokulardisplay. In der HoloLens-Applikation sind die Möglichkeiten von Sprachkommandos, Gestersteuerung und Augenblinzelsteuerung getestet worden. Letztere Steuerung ist in der HoloLens nicht möglich und wurde mithilfe eines Beobachters, der beim Blinzeln der Nutzer ein Bluetooth Gerät bedient, welches mit der HoloLens verbunden ist, simuliert. Bei der Evaluation der verschiedenen Interaktionsmethoden mit medizinischem Personal in einer Testumgebung stellte sich Augenblinzeln als vorteilhafteste Methode heraus. Diese Interaktionsweise ist vorteilhaft, da der Nutzer bei der Patientenbehandlung keine Sprachkommandos laut aussprechen, was Patienten missfallen könnte und der Arm nicht, wie bei der Gestersteuerung, durchgehend ausgestreckt sein muss, um interagieren zu können. Zudem zeigte sich, dass diese Methode die schnellste Methode für die Nutzer darstellt.

Die Entwicklung neuer Systeme, die dem medizinischen Personal helfen kann, geschieht kontinuierlich und wird durch neue Hardware für immer neue Anwendungsfälle erweitert. Besonders mit Blick auf die bisherigen Arbeiten, können weitere Software basierte Systeme in die bereits bestehenden Forschungsarbeiten integriert werden, um dem Personal noch bessere Hilfe im täglichen Umgang mit Patienten und Wunden an die Hand zu geben. Ein mögliches System wird in den folgenden Abschnitten beschrieben und mithilfe von ML implementiert.

# 4 Relevance

## 4.1 Aktueller Wundpflegeprozess

Um die Verbesserungsmöglichkeiten des Wunderkennungssystem einordnen zu können, muss zunächst der aktuelle Behandlungsprozess einer Wunde betrachtet werden. Dafür wurden bereits durchgeführte Ethnografien und Wundprozessmodellierung herangezogen, da die aktuelle epidemiologische Situation des Sars-Cov2 Virus eine der Arbeit vorangehende Befragung von medizinischem Personal als potenziell gefährlich und dem medizinischen Personal zusätzlich Belastung bereitend eingestuft wurde. Als Grundlage dient eine bereits ausgearbeitete Prozessmodellierung, die durch das BMBF-Projekt ARinFLEX basierend auf der Publikation von Klinker et al. [66] bereitgestellt wird. Diese zeigt den aktuellen Prozess bei der Wundbehandlung. Besonders mit Blick auf die Erstversorgung werden weitere Schritte, wie zum Beispiel die Erstbeurteilung und Abfrage von Patientenkonditionen, benötigt, um alle nötigen Informationen über eine Wunde zu sammeln. Bei der genauen Beurteilung einer Wunde gibt es zudem viele verschiedene Vorgehensweisen und festgehaltene Daten aus vielen verschiedenen Quellen [16] [17] [18] [67]. Daher wird die nachfolgende Beschreibung der Beurteilung besonders auf die Punkte eingehen die universell bei allen eingesehenen Quellen genannt werden.

Zunächst muss ein Pfleger die eigenen Vorbereitungen treffen. Diese bestehen aus der Desinfektion der Hände sowie dem Anziehen von sauberen Handschuhen. Bevor eine Wunde beurteilt werden kann, muss sie gereinigt werden, damit keine Fehler entstehen. Im Anschluss werden wichtige Beobachtungen festgehalten, die in Tabelle 4.1 aufgelistet und kurz beschrieben werden. Nachdem die Wunde beurteilt wurde, werden relevante Informationen über die medizinische Historie des Patienten eingeholt. Dabei wird besonders auf Vorerkrankungen wie Diabetes, Venen- und Arterieninsuffizienz geachtet, aber auch aktuelle Medikamenteninnahmen und andere mögliche relevante Daten erfasst. Wenn noch nicht während der Wundbeurteilung geschehen, werden einige Bilder der Wunde gemacht, um diese der Patientenakte beizulegen und später einen Referenzwert für den Heilungsprozess sowie Material für die Begutachtung durch anderes Personal zu haben. Im Anschluss wird die Wunde angemessen versorgt. Anschließend entsorgt die Pflegekraft die Handschuhe und desinfiziert die Hände wieder. Im letzten Schritt muss die Pflegekraft alle relevanten Daten entweder auf Papier oder an einem Computer dokumentieren, damit die Daten geltend sind und bei der nächsten Wunduntersuchung bereitstehen. Ein Problem, das während den Untersuchungen des Wundprozesses erkannt wurde, ist, dass das Pflegepersonal meist mit Tricks oder einer zweiten Person arbeiten muss, um die Dokumentation und Bildaufnahme effizient zu gestalten [66]. Denn die Menge an Daten, die während der Behandlung des

Tabelle 4.1: Tabelle mit wichtigen Metriken und Beobachtungen für die Wundbeurteilung.

Beobachtung	Erklärung
Wundphase	Aktuelle Phase der 4 möglichen Wundheilungsphasen (vgl. Unterabschnitt 3.1.2).
Wundbeläge	Beläge wie Nekrose, Fibrin oder Schorf.
Wundauswurf	Material- oder Flüssigkeitsauswurf
Wundgröße	Messung der Breite, und Höhe mit einem Lineal und falls möglich die Wundtiefe durch Tiefenindikatoren
Wundtyp	Diabetisches Fußsyndrom, venöser Ulkus, Dekubitus oder Andere
Schmerhaftigkeit	Keine Schmerzen bis starke Schmerzen
Infektion	Mögliche Infektion der Wunde
Wundränder	Charakteristika des Wundrands und der Wundumgebung

Patienten zusammenkommen und dokumentiert werden müssen, ist groß und schwer zu merken. Aus diesem Grund schreiben manche Pflegekräfte wichtige Werte entweder auf ein Papierlineal oder ihre Hände. Zudem ist es nur schwer möglich gleichzeitig mit einem Lineal eine Wunde zu messen und gleichzeitig gute Fotos mit einer Kamera zu machen. An dieser Stelle ist eine zweite Person hilfreich, die beide Aufgaben übernehmen kann. Zusätzlich bietet eine Kamera, die von Patient zu Patient getragen wird, ein mögliches Infektionsrisiko, weshalb diese auch desinfiziert werden muss, was je nach Modell sehr aufwändig sein kann. Insgesamt wurde der Prozess von den befragten Pflegekräften als ineffizient und hygienisch aufwändig eingestuft. [16] [17] [18] [66] [67]

## 4.2 Vergleich der Prozesse

Um den Prozess der Wundbehandlung effizienter zu gestalten, wird in dieser Arbeit ein Klassifizierungsprogramm für Wunden mit ML entwickelt. Damit ein neues Produkt sinnvoll in den Alltag einer Pflegekraft integriert werden kann, muss es einen deutlichen Mehrwert bieten und sollte bereits studierte Abläufe nicht unnötig unterbrechen [64] [68]. Da das Klassifikationsprogramm auf Bilder arbeitet, müssen die Bilder der zu klassifizierenden Wunde zuvor manuell aufgenommen werden. Da dies, wie zuvor beschrieben, im aktuellen Prozess mit normalen Kameras funktioniert, kann das Programm erst nachdem die Wundversorgung abgeschlossen und die Bilder an einem Computer hochgeladen wurden, genutzt werden. Dies würde einen zusätzlichen Schritt am Ende des komplexen Prozessablauf hinzufügen. Außerdem ist der Wundtyp zu diesem Zeitpunkt bereits festgestellt, was eine Klassifizierung obsolet macht. Da der aktuelle Ablauf der Wunddokumentation allerdings nicht optimal ist und an einigen Stellen in Begriff ist, mit neuen technischen Möglichkeiten unterstützt zu werden (vgl. Abschnitt 3.3, Abschnitt 3.4), wird der neue, beispielhafte Wundpflegeprozess auf der bestehenden Arbeit von Klinker et al. aufgesetzt. Eine Einpflegung des Programms als Subsystem in andere, bereits vorgestellte Systeme zur Wunddokumentation ist eben-

falls möglich. Damit Pflegekräfte keine Hilfe bei der Fotografie der Wunde und weniger Desinfektionen pro Patient benötigen, hat Klinker et al. einen Wundpflegeprozess mit der Augmented Reality Brille HoloLens modelliert (vgl. auch Abschnitt 3.4) [66]. Dieser Prozess, welcher in Abbildung 4.1 dargestellt ist, basiert auf der Nutzung einer Applikation auf der HoloLens, welche sowohl die Wundbeurteilung als auch Fotografie der Wunden ermöglicht. Dadurch sollen die aktuell verwendeten Tricks oder zusätzlich benötigten Personen bei der Wundbehandlung abgelöst und dem Pflegepersonal ein leichterer Ablauf an die Hand geben werden. Die in der Arbeit entwickelte Applikation bietet eine prototypische Version einer Dokumentationsumgebung. Patienten können ausgewählt, Informationen abgelesen und neue Wunden dokumentiert werden. Bei der Wunddokumentation wird zunächst ein Foto gemacht und dazu die Größe der Wunde angezeigt, sofern die Ankerpunkte für die Messung gesetzt wurden. Im Anschluss können verschiedene Daten der Wundbeurteilung, wie zum Beispiel betroffene Gewebestrukturen oder Wundzustände, ausgewählt werden. In einer vollständigen Applikation wären mehr einzutragende Daten und Funktionen möglich.

Nachdem der Erläuterung des Prozesses von Klinker et al., wird nachfolgend der Einsatz des Klassifizierungsprogramms in diesen Prozess beschrieben, wie in Abbildung 4.2 visualisiert. Das vorgeschlagene Programm verändert nur in den drei Schritten 'Wundgröße mit der HoloLens messen', 'Foto per Sprachbefehl mit der HoloLens aufnehmen' und 'Wunde dokumentieren' ab dem Zeitpunkt des Start des Wunddokumentation einige Teilschritte. Nachdem die Wunddokumentation beginnt, wählt die Pflegekraft die Ankerpunkte für die Größenmessung der Wunde aus. Im Anschluss schießt sie mit der HoloLens ein Foto und prüft ob es den geltenden Vorschriften entspricht. Nachdem das Foto als gültig übernommen wird, sendet die HoloLens das Bild an einen Server, zum Beispiel im lokalen Netzwerk, auf welchem das ML Modell aufgesetzt ist. Nach Eingang des Bildes wird die Wunde klassifiziert und das Ergebnis zurück an die HoloLens gesendet. Während die Wunde klassifiziert wird, kann die Pflegekraft bereits einige Daten über die Wunde eintragen. Wenn das Klassifizierungsergebnis bei der HoloLens ankommt, füllt die Dokumentationsumgebung den Wundtyp in den Behandlungsbogen ein. Die Pflegekraft kann währenddessen die restlichen Daten eintragen und überprüft zum Schluss noch einmal das Ergebnis der Wunderkennung. Danach bestätigt die Pflegekraft die Dokumentation, um den Prozess abzuschließen. Diese vorgestellte Erweiterung des Arbeitsprozesses wird vor allem automatisch durch die entsprechende Applikation ausgeführt. Die Pflegekraft muss nur vor dem Abschluss der Dokumentation die Daten überprüfen, damit kein Fehler entsteht. Mit dem richtigen HoloLens-Interface lässt sich so ein Mehraufwand verhindern und stattdessen eine leichtere Erkennung eines Wundtyps implementieren. Das auf dem Server laufende Programm kann durch eine minimale Abwandlung des in Kapitel 5 vorgestellten serverseitigen Systems implementiert und für verschiedene Bedürfnisse angepasst werden.

Damit das verwendete ML Modell auch korrekt funktionieren kann, muss es zunächst mit den richtigen Daten trainiert werden. Da in dieser Arbeit überwachtes ML eingesetzt wird, müssen neben den Bildern auch die Lokalisierungen und Klassifizierungen der Wunden

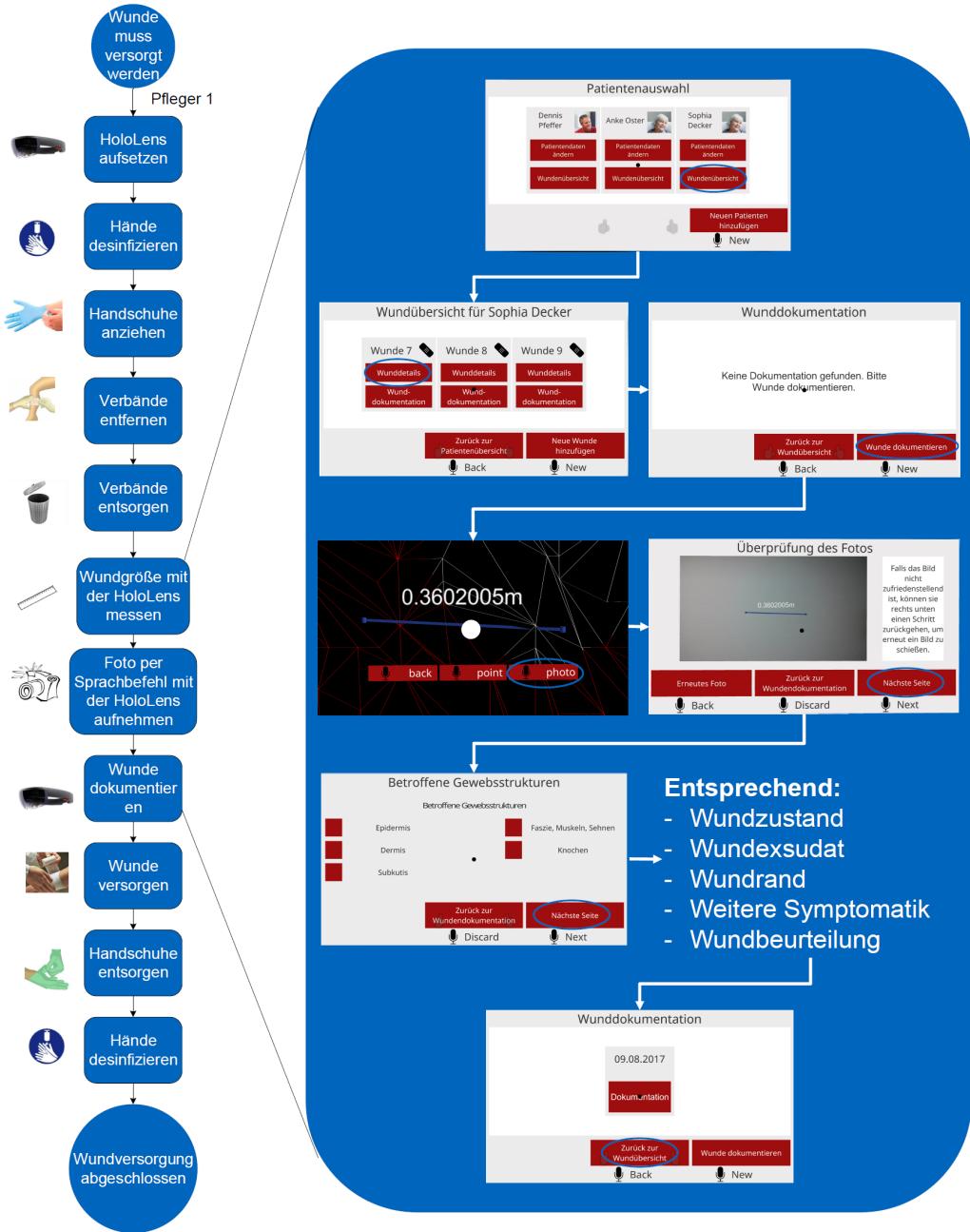


Abbildung 4.1: Wundpflegeprozess mit Augmented Reality Brille nach Klinker et al. [66]

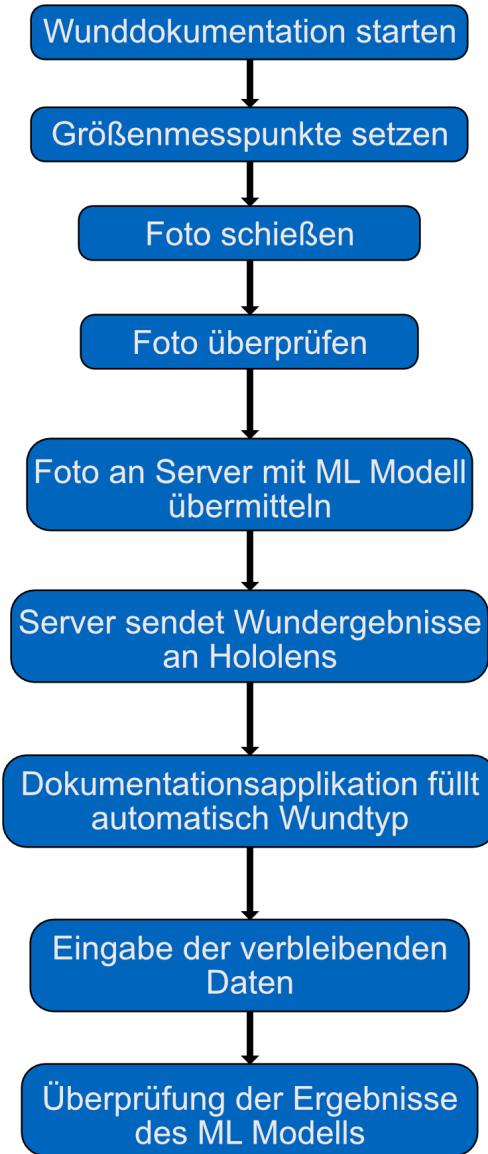


Abbildung 4.2: Um ML gestützte Wundklassifizierung erweiterter Prozessablauf für die Schritte 'Wundgröße mit der HoloLens aufnehmen' bis 'Wunde dokumentieren' aus Abbildung 4.1.

in das Modell eingegeben werden (vgl. Abschnitt 3.2.1). Um die höchstmögliche Korrektheit zu garantieren, kann das Wissen von Pflegepersonal genutzt werden, um Wunden zu klassifizieren. Dafür wird ein zweiter Prozess, welcher vom Ersten zeitlich und räumlich getrennt ist, modelliert werden. Da in einer Vielzahl von Wundbildern nur eine Wunde zu sehen ist, kann das medizinische Personal mehr Bilder in kurzer Zeit annotieren. Zu Beginn setzt sich die Pflegekraft an einen Computer der Einrichtung und ruft die Annotationswebseite auf. Dort sind Instruktionen beschrieben, wie die Webseite funktioniert und was zu tun ist. Die Pflegekraft lädt daraufhin ein Wundbild hoch. Hat die medizinische Einrichtung bereits ein trainiertes ML Modell, so wird das Wundbild zunächst mithilfe dieses Modells klassifiziert. Wenn kein ML Modell vorhanden ist, wird die Pflegekraft direkt zur Annotationsoberfläche weitergeleitet. In dieser können neue BB platziert werden, Wundtypen der BB angepasst werden und von einem Modell erkannte Klassifizierungen korrigiert werden. Wenn die Pflegekraft alle Wunden korrekt annotiert hat, schick sie die Daten an den Server und kann sich entscheiden ein weiteres Bild hochzuladen oder die Webseite zu schließen. Der aufgezeigte

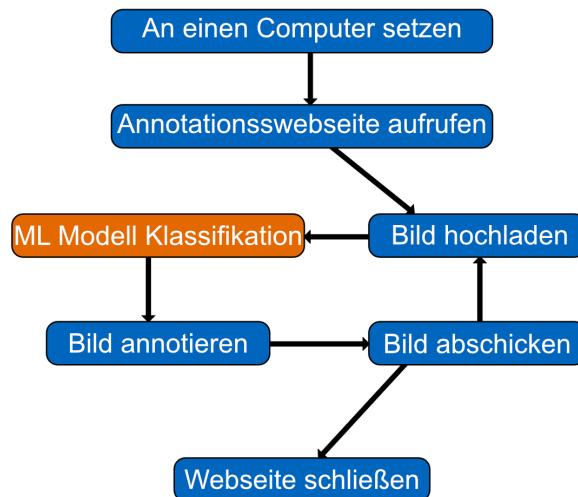


Abbildung 4.3: Annotationsprozess von Wundbildern durch eine Pflegekraft. Orange markierter Schritt wird übersprungen, falls noch kein Wunden erkennendes ML Modell vorhanden ist.

Prozess stellt die einfachste Möglichkeit zur Annotation der Daten dar. Komplexere Systeme über automatische Bildweiterleitungen und Erkennungen nach dem Fotografieren der Bilder sind ebenfalls möglich, erfordern aber zusätzliche Implementierungen auf der Softwareseite, welche nicht Teil dieser Arbeit sind. Wichtig bei diesem Prozess ist die einfache und schnelle Nutzung der Webseite. Dadurch wird das Pflegepersonal bei einigen Bildern pro Tag nicht stark belastet. Die Annotation von Bildern kann zu jedem Zeitpunkt am Arbeitstag geschehen sobald Bilder aus den alltäglichen Behandlungen für solch eine Nutzung bereitstehen.

Mit diesem Prozess wird eine minimale Anzahl an Aufgaben zu den bereits vorhandenen Prozesskonzepten mit Augmented Reality Brillen oder anderen Wunddokumentationssystemen

hinzugefügt, um die im nächsten Abschnitt beschriebenen Verbesserungen zu maximieren.

### 4.3 Potenzial und Verbesserungen durch das Programm

Besonders mit Blick auf die bereits stark ausgelasteten und sehr eingearbeiteten Pflegekräfte, müssen Prozessänderungen große Vorteile bieten und sinnvoll in den Arbeitsalltag integrierbar sein, um das Personal dafür zu gewinnen [64] [68]. Besonders mit Blick auf die Literatur, die für die Erklärungen der Wunden in Abschnitt 3.1 genutzt wurde, haben sich immer wieder Details bei den Wunden unterschieden. Vor allem tritt dies in Form fehlender Informationen oder nicht weiter spezifizierten Daten auf. Dies kann zu unterschiedlichem Verständnis von Wunden führen und besonders bei neuen Pflegekräften, die noch im Lernprozess sind, für Verwirrung sorgen. Zusätzlich sind dies nur Dokumente aus dem deutschen und englischsprachigen Raum. Mit Blick auf internationale medizinische Versorgung, kann noch eine stärkere Diskrepanz zwischen Wunddefinitionen existieren. Aus diesen Grund bietet ein automatisiertes Wunderkennungssystem die Möglichkeit dies zu vereinheitlichen. Auch wenn zunächst eine große Datenbank angelegt werden muss, um ein ML Modell für medizinischen Einrichtungen zu generieren, kommt am Ende durch viele verschiedene Bilder ein Modell heraus, dass Bilder immer gleich bearbeitet und die gleichen Wunden darauf erkennt. Eine zugrundeliegende Voraussetzung ist eine hohe Korrektheit des ML Modells, um so einen tatsächlichen Mehrwert zu bieten.

Mit dem Blick auf den sehr aufwändigen Wunddokumentationsprozess lassen sich viele automatische Systeme entwickeln, um diesen zu vereinfachen. Durch die Arbeit von Wang et al. [64] wurde gezeigt, wie ein digitaler Dokumentationsbogen mit einer einhergehenden Datenorganisation die Arbeitsschritte deutlich beschleunigen kann. Dies ist auch mit entsprechender Routine und ausgearbeiteten Applikationen mit einer Augmented Reality, wie als Konzeptversion von Klinker et al. [66] gezeigt, möglich. Damit solche Systeme eine noch höheren Mehrwert bieten, müssen intelligente Systeme als neue Komponenten eingebaut werden können. Ein solches System ist die ML gestützte Wundklassifizierung, welches einen weiteren Schritt der Dokumentation übernehmen kann, während die Pflegekraft sich um den Patienten kümmert.

Auch mit Blick auf die aktuell sehr schlechte Datenlage in Bezug auf Wundbilder und korrekte Wundannotationen hat die Annotationswebseite das Potenzial eine aktuell noch vorhandene Lücke für die Wissenschaft zu füllen. Denn bisher werden in den meisten Arbeiten einzelne Experten im Bereich von Wunden konsultiert, um die für die spezifische Arbeit gesammelten Bilder zu markieren [52] [60] [69]. Eine Datenbank, die ständig mit neuen Bildern versorgt wird, bietet großes Potenzial, um die Wissenschaft im Allgemeinen einfacher zu gestalten. Zudem ist das Programm nicht auf Wundbilder beschränkt, sondern kann für beliebe Bereiche, die annotierte Bilddaten benötigen, angepasst werden. Auch Anpassungen der BBformen ist möglich. Denn viele Arbeiten verwenden Polygone hohen Grades, um eine deutliche genauere Flächenangabe zu erhalten [52] [69] [60].

# 5 Design Cycles

## 5.1 Design Cycle 1

Der erste Design Cycle beschreibt die an diese Arbeit vorangegangene Arbeit am Lehrstuhl für Wirtschaftsinformatik an der Technischen Universität München. Dabei wurde zunächst ein Proof of Concept System entwickelt, um die Funktionalität des Systems zu überprüfen. Um eine ausreichende Funktionalität für eine weitere Forschungsarbeit zu erreichen, muss der Prototyp zwei wichtige Funktionen erfüllen. Zum einen muss das ML Modell Wunden und Teile von Wunden korrekt erkennen können und zum anderen muss es ein nutzerfreundliches System zur Klassifizierung von Bildern für das Training eines ML Modells geben.

### 5.1.1 Grundlagen und Implementierung

Aufgrund der beiden Anforderungen an den Prototyp wurden mehrere Implementationsmöglichkeiten untersucht. Für das ML Modell bietet sich das freie und quelloffene ML Framework Tensorflow [10] von Google besonders durch ein schnelles Aufsetzen der Software und ein einfaches Training von ML Modellen an. Tensorflow wurde als GPU basiertes Modell mit 253 Trainingsbildern, sowie 59 Testbildern trainiert. Das Aufsetzen des Projekts wurde entsprechend des Tutorials von Edje Electronics durchgeführt [70]. Zudem werden auch Teile seines Codes für die Initialisierung des Klassifizierungsprozesses in der Datei genutzt, welche sich im Original in seinem Repository finden, und in teils angepasster Form im Repository dieser Arbeit [71]. Die Trainingsbilder wurden aus den Suchergebnissen der Google Bildersuche heruntergeladen und im Anschluss manuell mit, soweit wie möglich, korrekten Begrenzungsboxen und Labels versehen. Das Training wurde bis zu einem konstanten Verlauf der Loss Funktion durchgeführt. Um eine möglichst gute Aussage über die Funktionalität des Systems treffen zu können, kann das Modell zwei Klassen unterscheiden. Zum einen Ulzera und zum anderen nekrotisches Gewebe. Der besondere Vorteil dieser beiden Kategorien ist die einfache Identifizierung, da keine medizinische Ausbildung nötig war, um diese Klassen zu annotieren.

Das Ziel des Annotationsprogramms ist es eine möglichst einfache, überall verfügbare Lösung bereitzustellen, mit der medizinisch ausgebildetes Personal Wundbilder hochladen und annotieren kann. Durch dieses Programm kann dann eine große, medizinisch korrekt klassifizierte Datenbank von Wundbildern angelegt werden, um ML Modelle zu trainieren und auch andere Forschungsprojekte zu unterstützen. Nach der Besprechung einiger Lösungsansätze mit meinem Betreuer Kai Klinker, welcher die Leitung des Projektes innehat, ist ein Browserbasiertes Programm als beste Lösung hervorgetreten. Dieses Programm ist in Form einer Webseite realisiert, die als Serverinfrastruktur eine NodeJS Implementierung in Kombination

mit Express benutzt. Der NodeJS Server implementiert dabei einfache Webserver Funktionen und erlaubt den Aufruf des Python basierten Tensorflow Frameworks. Nachdem Nutzer ein Bild hochgeladen haben, müssen diese 10 bis 20 Sekunden warten bis das ML Modell die Klassifizierung durchgeführt hat und die Daten an die Nutzer zurückschickt, damit diese die Annotation beginnen können. An dieser Stelle ist es wichtig festzuhalten, dass das Tensorflow Framework keine Parallelität unterstützt. Wenn also ein Nutzer ein Bild hochlädt und ein anderer Nutzer ein zweites Bild hochlädt, so muss der zweite Nutzer warten bis das ML Modell mit dem ersten Bild fertig ist, um das zweite Bild zu bearbeiten. Dadurch entsteht bei viele parallelen Nutzern eine potenziell lange Warteschlange. Die Optimierung dieses Problems wird in dieser Arbeit nicht weiter behandelt, da die Implementierung einer solchen Parallelität tief in das Framework vordringen würde und zudem die zugrundeliegende Machbarkeit in Frage steht. Das Serverprogramm hat sich in den nachfolgenden Design Cycles nur noch minimal verändert, um benötigte Parameter für die HTML Webseite bereitzustellen. Daher wird dieses Programm hier nicht weiter besprochen, kann aber über das Coderepository abgerufen werden [71]. Der Nutzer erhält auf der Webseite alle nötigen Instruktionen und die Möglichkeit ein Bild hochzuladen (vgl. Abbildung 5.1). Nachdem der Server das Bild empfangen hat, gibt er das Bild an das Tensorflow-Modell weiter und lässt mögliche Wunden und nekrotisches Gewebe erkennen. Im Anschluss erhält der Nutzer die Eingabeoberfläche mit den erkannten Begrenzungsboxen und Labels (vgl. Abbildung 5.2). Diese müssen vom Nutzer gegebenenfalls in Position, Größe oder Label angepasst werden. Im Fall von nicht erkannten Wunden, kann der Nutzer diese hinzufügen, um eine korrekte Klassifizierung an den Server zu senden. Dieser speichert die Daten in dem für Tensorflow benötigten Format zusätzlich zu dem Bild ab, sodass ein weiteres Training des Modells möglich ist. Dem Nutzer stehen die Funktionen in Tabelle 5.1, welche in dieser Version noch nicht übersetzt wurden, zur Verfügung.



Abbildung 5.1: Startseite der Klassifizierungswebseite. (Design Cycle 1)

Tabelle 5.1: Tabelle mit allen Funktionen des Klassifizierungsprogramms. (Design Cycle 1)

Funktion	Erklärung
Select	Auswählen einer BB durch klicken in die Fläche
Move	Verschieben der aktuell ausgewählten BB durch gedrücktes Ziehen der Maus
Add	Hinzufügen einer neuen BB durch Klicken und anschließendes Ziehen mit gedrückter Maus
Delete	Löschen der aktuell ausgewählten BB
Clear	Löschen von allen BB
Move	Verschieben der aktuell ausgewählten BB
Tag	Ändern des Labels der ausgewählten BB
Resize	Verändern der Größe der ausgewählten BB durch Klick in eine der beiden angezeigten Ecken und anschließendes Ziehen mit gedrückter Maus
Submit	Senden der BB an den Server

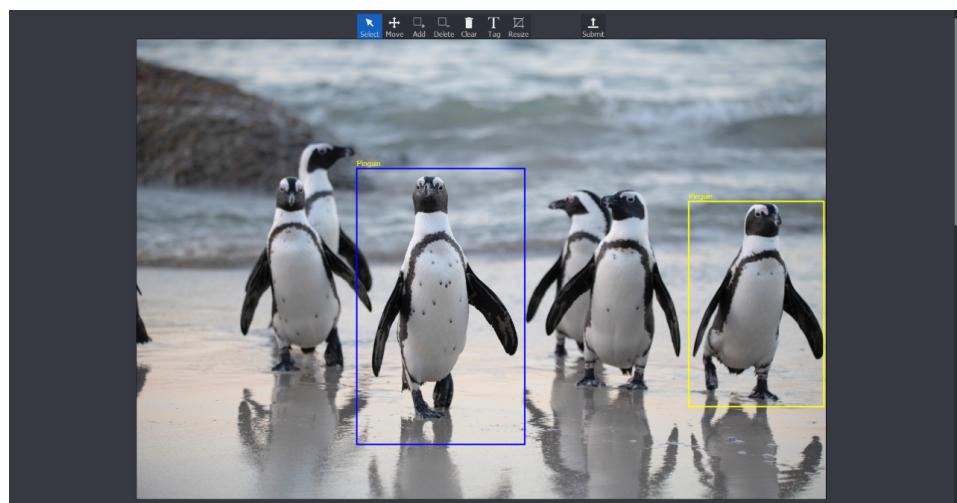


Abbildung 5.2: Eingabeoberfläche des Klassifizierungsprogramms, um Begrenzungsboxen und Labels zu bearbeiten. (Design Cycle 1)

### 5.1.2 Nutzertests

Die Probanden haben die Software mit 5 Testbildern von Google Images zugesandt bekommen. Zunächst mussten alle Testpersonen ein Bild nach dem anderen hochladen. Sie wurden gebeten alle nicht korrekt platzierten oder benannten Begrenzungsboxen anzupassen oder zu löschen und im Anschluss die jeweils verbleibenden Objekte zu markieren. Auf dem ersten Bild waren Gesichter von Personen zu markieren. Im Anschluss mussten alle Autos auf einer Straße markiert werden. Im darauffolgenden Bild wurden alle Katzen markiert. Das vierte Bild zeigt eine Strand, an welchem alle Strandkörbe markiert werden musste, bei denen die Zahl auf der Rückseite erkennbar war. Im letzten Bild mussten alle Pinguine erfasst werden. Diese Bilder wurden ausgewählt, um verschiedene Situationen im Programm hervorzurufen, wie sie auch bei einer Wundklassifizierung vorkommen könnten. Zum einen wurden einfach erkennbare Wunden simuliert (1, 3). Des Weiteren wurde mit Objekten, die nicht parallel zu den Bildschirmachsen sind, ein schiefes Bild simuliert (2, 4). Ebenfalls wurden Klassifizierungen getestet, die mehrere überlappende Begrenzungsboxen benötigen (3, 5). Es gab bei diesem Test keine Aufgabenstellung das spezifische Features genutzt werden müssen.

Das Tensorflow Modell wurde mit einigen Testbildern, die ebenfalls von Google Images stammen, getestet, um zu prüfen ob das System überhaupt auf Wundbilder in verschiedenen Lichtsituationen reagiert. Dafür wurden die Bilder an das Modell zur Klassifizierung übergeben und im Anschluss visuell überprüft, ob eine sinnvolle Lokalisierung und Klassifizierung vorhanden war.

### 5.1.3 Resultate

Zunächst werden die aufgetretenen Probleme der Nutzer beschrieben und daraufhin welche Nutzungsverhalten die Testpersonen zeigten. Während des Tests von allen Nutzern ist die Situation aufgetreten, dass die Funktionsweise des Löschens nicht klar war und eine falsche Begrenzungsbox gelöscht wurde. Die Nutzer erwarteten nach dem Klicken des entsprechenden UI Elements auf die zu löschen Begrenzungsbox klicken zu müssen. Stattdessen wird die aktuell ausgewählte Begrenzungsbox gelöscht. Zudem war die Notwendigkeit der Auswahl der zu bearbeitenden Begrenzungsbox nicht intuitiv, da nicht klargestellt wurde welche Funktionen mit der ausgewählten Begrenzungsbox interagieren. Es gab bei zwei Nutzern das Problem, dass die Funktionsweise des Bildhochladens nicht klar war und erst nach einigen Versuchen korrekt ausgeführt wurde. Ein weiteres Problem war die Auswahl von bestehenden Begrenzungsboxen, da ein Klicken auf den Rand der Begrenzungsbox nicht dazu geführt hat, dass sie ausgewählt wurde. Auf einigen Bildern konnten die Probanden die Labels der Begrenzungsboxen nicht lesen, da der Text in Gelb dargestellt wird und der Hintergrund ebenfalls eine sehr helle Farbe hatte. Ein weiteres Problem bei der Auswahl von Begrenzungsboxen sind überlappende Flächen. Dabei kann immer nur eine Begrenzungsbox ausgewählt werden. Sollte nun eine Begrenzungsbox komplett in einer anderen liegen, so muss, abhängig von der Einfügereihenfolge, die äußere Begrenzungsbox zur Seite geschoben

werden, um die innere Begrenzungsbox auswählen zu können.

Beim Hinzufügen neuer Begrenzungsbox bevorzugten Nutzer es direkt abzubrechen, statt die Begrenzungsbox anschließend in die korrekte Position und Größe zu bringen. Die generierten Begrenzungsboxen wurden immer gelöscht, wobei nur ein Nutzer die Funktion alles auf einmal zu löschen nutzte. Nach dem Löschen wurden für alle Objekte neue Begrenzungsboxen erstellt. Die Nutzer, welche bereits platzierte Begrenzungsboxen mit der Größenveränderung angepasst haben, hatten Probleme die beiden Ecken, an denen die Größenänderung möglich ist, zu identifizieren. Zum einen, weil die gelben Indikatoren unter der eigentlich Begrenzungsbox gezeichnet werden, und zum anderen, weil die gelbe Farbe auf einigen Hintergründen schwer zu erkennen war.

Bei einem Nutzer, welcher eine veränderte Fenstergröße für den Browser benutze, haben sich UI Elemente ungünstig verschoben, sodass sie nicht mehr auswählbar waren. Dies führte dazu, dass die Fenstergröße geändert und das Bild erneut hochgeladen werden musste.

Das Prototyp Tensorflow Modell ist in der Lage Nekrose und Ulzera grundsätzlich zu richtig zu erkennen. Dies ist die Grundvoraussetzung für das Training von tatsächlichen Wundtypen. Dies wird in Abschnitt 5.4 und Abschnitt 5.5 mithilfe der Medetec Datenbank [22] realisiert.

#### 5.1.4 Gelerntes

Besonders wichtig ist es dem Nutzer Informationen über die Funktionsweise des Programms an die Hand zu geben. Damit lässt sich der Lernprozess deutlich abkürzen, da das Programm nur wenige Funktionen anbietet und den Nutzer nicht mit vielen Erklärungen überfordert. Zudem ist es wichtig die Interaktionsmöglichkeiten an bereits sehr bekannte Programme anzulehnen. Dies ist besonders bei Veränderung der Größe von Begrenzungsboxen sichtbar, da bekannte und viel genutzte Bildbearbeitungsprogramme wie Photoshop oder Affinity Photo deutlich mehr Optionen zur Größenanpassung bieten. Eine weitere Funktion, welche die zuvor genannten Programme implementieren, ist eine Hierarchie zur Auswahl von Objekten. Dies ist hilfreich, um überlappende Objekte zielgerichtet auswählen zu können. Aufgrund von unterschiedlichen Bildschirmgrößen und Fenstergrößen ist zudem ein sogenanntes Responsive Design notwendig, damit die UI Elemente immer korrekt dargestellt werden und benutzbar bleiben.

## 5.2 Design Cycle 2

### 5.2.1 Grundlagen und Implementierung

Basierend auf den Ergebnissen aus Abschnitt 5.1 wurde das web-basierte Artefakt weiterentwickelt, um die Probleme zu entfernen und nützliche Funktionen hinzuzufügen. Das erste Element ist ein Hilfe Fenster, welches über die Eingabeoberfläche aufgerufen werden kann,

um die Funktionsweise aller UI Elemente einzusehen (vgl. Abbildung 5.4). Dieses Fenster wird bei erstmaliger Nutzung des Programms nach der Weiterleitung angezeigt und bei weiterer Nutzung durch das Setzen eines Cookies nur noch nach dem Klicken des Hilfe Knopfs angezeigt. Zudem wurden alle UI Elemente, die Text enthielten, in die deutsche Sprache übersetzt, da anzunehmen ist, dass das komplette medizinisches Personal im deutschen Gesundheitssystem Deutsch beherrscht. Wie in Tabelle 5.1 beschrieben, arbeiten einige Funktionen auf der aktuell ausgewählten Begrenzungsbox, während zum Beispiel 'Clear' keine ausgewählte Begrenzungsbox benötigt. Um dies besser hervorzuheben, wurden die UI Knöpfe, welche eine Auswahl benötigen, von den Anderen räumlich getrennt. Dies soll dem Nutzer, neben den Erklärungen der Funktionen, verdeutlichen, dass unterschiedliche Funktionsweisen verwendet werden. Des Weiteren wird, um die Labels leserlicher zu gestalten, eine schwarze Umrandung um den gelben Text gezeichnet, sodass diese unabhängig vom Hintergrund immer erkennbar ist.

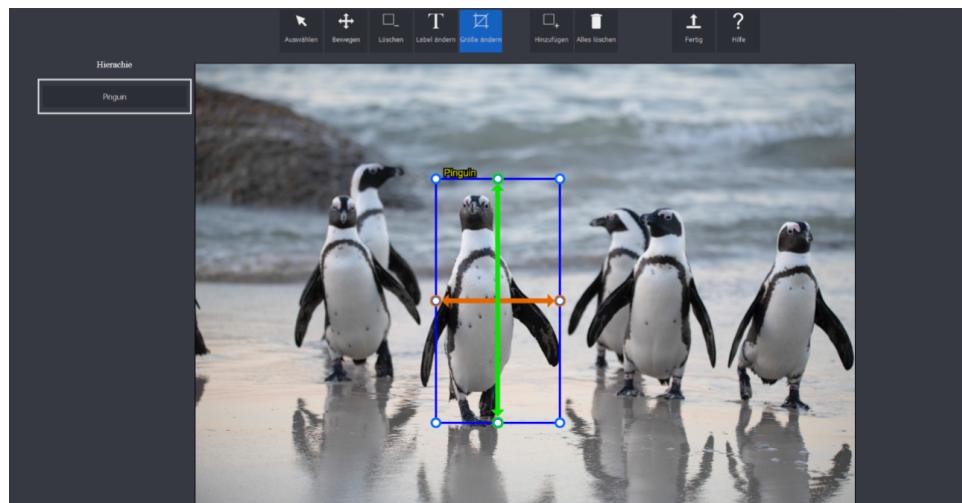


Abbildung 5.3: Eingabeoberfläche des Klassifizierungsprogramms, um Begrenzungsboxen und Labels zu bearbeiten, sowie Achsen, an denen über die zentralen Punkte die Größe der BB verändert werden kann. (Design Cycle 2)

Um die Nutzung der Größenveränderung attraktiver zu gestalten, wurde ein Format wie in Affinity Photo [72] implementiert. Dabei kann jede Begrenzungsbox nun an acht Punkten vergrößert oder verkleinert werden. Vier Punkte sind an den Ecken platziert, so dass man nun jede Ecke mit den zwei angrenzenden Kanten beliebig verschieben kann. Die anderen vier Punkte sind jeweils im Zentrum der Kanten platziert und erlauben eine Größenänderung nur entlang der dazu senkrechten Koordinatenachsen (vgl. Abbildung 5.3). In Abbildung 5.3 ist am linken Rand des Fensters die zuvor als nützlich identifizierte Hierarchie zu erkennen. Diese erlaubt es dem Nutzer durch einen Klick auf das entsprechende Element auch Begrenzungsboxen auszuwählen. Dadurch sind verschachtelte Begrenzungsboxen auswählbar,

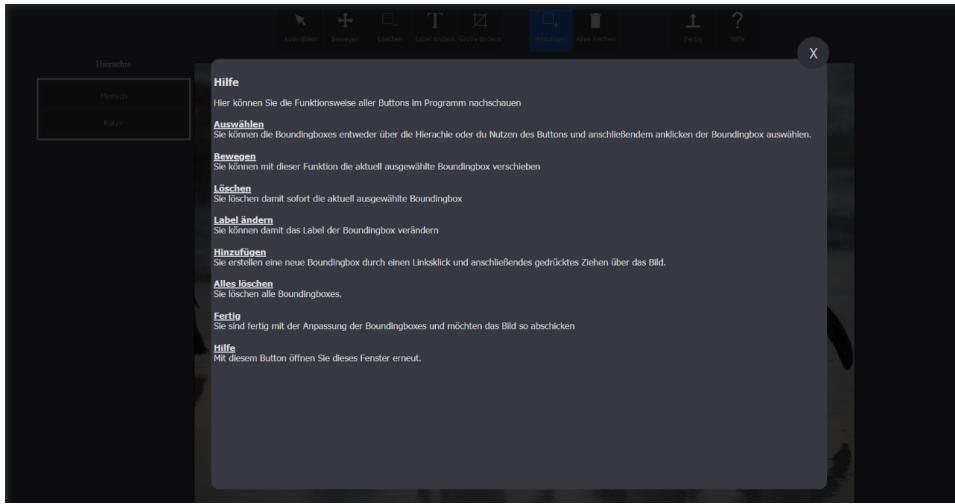


Abbildung 5.4: Hilfefenster mit Erklärungen der Funktionsweise des Programms. (Design Cycle 2)

ohne die Äußenen verschieben zu müssen. Zudem lassen sich alle Tags auch außerhalb des eigentlichen Bildes ablesen, um die Korrektheit zu überprüfen.

Um unterschiedlichen Fenster und Bildschirmgrößen entgegen zu wirken, wurde ein Responsive Designs implementiert, welches bei bestimmten Größen des Browserfensters die UI Elemente herunter skaliert. Dadurch sollen immer alle Elemente sichtbar und benutzbar bleiben. Um diese Funktion zu unterstützen und die Hierarchie gut lesbar zu behalten, wurde eine Skalierung für die hochgeladenen Bilder eingefügt. Bilder die größer als das Fenster sind, werden nun abhängig vom Seitenverhältnis so skaliert, dass sie immer in das Fenster passen.

### 5.2.2 Nutzertests

Es ist anzumerken, dass ein Nutzer aus zeitlichen Gründen nicht an dem Design Cycle teilnehmen kann, da er geschäftlich einige Wochen in eine andere Stadt musste. Der Nutzer wurde nicht durch einen neuen Probanden ersetzt, da im ersten Design Cycle festgestellt wurde, dass bei fünf Nutzern bereits die meisten Probleme identifiziert wurden. Die Testumgebung ist identisch mit der Testumgebung aus Abschnitt 5.1. Die Nutzer haben das Programm als Ganzes erhalten und wurde gebeten verschiedene Dinge auf neuen Bildern zu markieren. Zusätzlich sollten die Nutzer die neuen Funktionen des Programms in diese Aufgaben in mindestens zwei Bildern mit einbeziehen. Dies bedeutet, dass zum Auswählen von Begrenzungsboxen die Hierarchie mitbenutzt werden sollte und statt zu kleine oder zu große Bilder zu löschen, diese auch mit der Größenveränderung angepasst werden sollen. Auf den Testbildern sind verschiedene Situationen dargestellt, um mögliche Programmsituationen, wie Überlappung und hohe Genauigkeit, zu simulieren. Auf dem ersten Bild mussten alle Tiere markiert werden. Das zweite Bild zeigt ein Fußballspiel, auf welchem alle Spieler markiert werden sollten. Das dritte Bild zeigt drei Dinosaurier, auf denen alle Dinosaurier

markiert werden sollten. Im vierten Bild waren alle Rennradfahnen zu markieren. Im letzten Bild mussten alle Pinguine markiert werden.

### 5.2.3 Resultate

Im zweiten Design Cycle haben die Nutzer ein gutes Verständnis der Funktionsweisen des Programms gezeigt. Es gab keine Probleme bei der aktiven Nutzung des Programms. Die aufgetretenen Probleme waren technischer Natur, die auf die unterschiedlichen Bildschirmgrößen der Nutzer, sowie eine fehlerhafte Gestaltung des Layouts zurückzuführen sind. Das Hauptproblem stellt das erste Bild dar. Dieses hat keine üblichen Pixelmaße, sondern 1110 Pixel in der Breite und 440 Pixel in der Höhe. Dies führt dazu, dass die Skalierung des Bildes nicht benötigt wurde, da das Bild in das Browserfenster passt. Dadurch wurde das Bild auf Bildschirmen mit niedrigerer Auflösung von der Hierarchie überdeckt. Dies führte zu einer schlechten Bedienung des Programms im linken Teil des Bildes. Ein weiteres Problem war die fehlende Aktualisierung der Hierarchie bei der Änderung eines Labels einer bestehenden BB. Es wurde festgestellt, dass die Position des Hilfe-Knopfs neben dem Sendeknopf ohne einen Sicherheitsmechanismus ungünstig für Nutzer sein kann. Ein Nutzer wollte das Hilfsfenster öffnen, hat dabei aber unabsichtlich das Bild abgesendet, ohne mit der Aufgabe fertig zu sein. Ein weiterer Fehler, der nur bei einem Probanden auftrat, ist die Möglichkeit beim gedrückten Ziehen auf dem Bild, das Bild mitziehen zu können. Dies passierte nur im Browser Edge von Microsoft.

Die Probanden haben das Programm mit mehr Selbstbewusstsein als zuvor genutzt, da die Grundfunktionen aus dem vorangegangenen Design Cycle bereits bekannt waren. Nach dem Hinweis, dass die neuen Funktionen des Programms mindestens zweimal mit in die Nutzung einbezogen werden sollen, haben vier von fünf Nutzern bei den ersten beiden Bildern die neuen Funktionen ausprobiert und haben danach das gleiche Nutzungsverhalten wie im vorangegangenen Design Cycle gezeigt. Die neuen Funktionen wurden durch die Erklärungen im Hilfsfenster bereits im vornherein erklärt und wurden dadurch ohne Verständnisprobleme benutzt. Die Probanden hatten nach den Aufgaben die Möglichkeit Feedback zu äußern. Dabei wurde nur angemerkt, dass ein Mechanismus für die Bestätigung des Absendens vorteilhaft sein würde.

### 5.2.4 Gelerntes

Die neuen Funktionen wurden positiv aufgenommen und machten die Bedienung deutlich leichter. Gleichzeitig haben die Probanden ihren zuvor gelernten Nutzungsablauf bevorzugt, um die Aufgaben zu erfüllen. Ein gesondertes Fenster, welches alle Erklärungen enthält, ist eine sinnvolle Ergänzung und sollte zu jedem Zeitpunkt aufrufbar sein. Des Weiteren ist es wichtig darauf zu achten, dass genug Bildschirmauflösungen und Bildschirmformate unterstützt werden, um zu verhindern, dass das Programm unbenutzbar wird. Mit den Ergebnissen des Nutzerverhaltens zeigt sich, dass das Programm für die Probanden beim zweiten Test sehr leicht zu benutzen ist, auch als neue Funktionen dazugekommen sind.

## 5.3 Design Cycle 3

### 5.3.1 Grundlagen und Implementierung

Im dritten Design Cycle wurden Verbesserungen des Programmverhaltens implementiert. Besonders Augenmerk lag darauf, dass alle Bilderformate auf allen gängigen Bildschirmauflösungen benutzbar sind. Dafür wurde das grundlegende UI Layout überarbeitet, um die Hierarchie und Bildfläche demselben Elternobjekt zuzuordnen. Dies führt mit entsprechenden Änderungen in den Eigenschaften der HTML Objekte dazu, dass beide Elemente immer nebeneinander angezeigt werden und sich nicht überlagern können. Des Weiteren wurden Fehler, die während der Implementierung der neuen Funktionalitäten im Programm auftauchten, behoben. Zu diesen zählten Ladefehler im UI und in der internen Verarbeitung, wodurch zum Beispiel Hierarchieelemente ihren Text nicht ändern konnten. Des Weiteren wurde eine Schaltfläche eingebaut, welche nach dem klicken des Sendeknopfes fragt, ob diese Aktion gewollt ist. Dies soll irrtümliches Einreichen von Bildern verhindern, sodass die Bilddatenbank nicht im Nachhinein manuell editiert werden muss, um möglicherweise falsche Bilder zu löschen.

### 5.3.2 Nutzertests

Für die Nutzertests wurden wie in den beiden vorhergehenden Design Cycles fünf verschiedene Bilder ausgesucht, die verschiedene Situationen simulieren sollen. Im ersten Bild mussten die Probanden alle Windräder markieren. Dieses Bild war zudem in einem Hochformat, um mögliche gedrehte Bilder abzubilden. Daraufhin mussten mehrere Personen in einem Raum markiert werden. Im Anschluss wurde dasselbe Bild von Hühner aus dem vorangegangenen Design Cycle verwendet, um die behobene Überlappung der Hierarchie zu testen. Daraufhin mussten die Tester einige Quallen in einem Bild annotieren. Im letzten Bild mussten aller Töpfe markiert werden. Zusätzlich wurden zwei weitere Personen mit medizinischen Bezug, eine Krankenschwester und eine Unfallbearbeiterin bei einer Berufsgenossenschaft, mit tatsächlichen Wundbildern und angeschlossenem ML System gebeten, das System wie im modellierten Alltagsprozess zu benutzen. Dabei wurde der erste der zwei zusätzlichen Tests von Philipp Reindl-Spanner [73] mit meinen Anweisungen durchgeführt. Nachdem ein Proband alle Bilder markiert hat, wurde er gebeten den Fragebogen zur System-Gebrauchstauglichkeit [74] auszufüllen. Dieser stellt eine Übersetzung des standardisierten und weit verbreiteten Fragebogens von Brooke et al. zur Befragung der Nutzerzufriedenheit dar [75]. In diesem werden dem Nutzer Zehn Fragen vorgelegt. Die Fragen können dabei von Null, gar nicht zutreffend, bis Fünf, stark zutreffend beantwortet werden. Nachdem alle Testpersonen den Fragebogen ausgefüllt haben, wird das Punktsystem des Fragebogens angewendet, um eine Gesamtpunktzahl auf einer Skala zwischen Null und 100 zu erhalten. Die Ergebnisse werden in dieser Arbeit mithilfe der Scoring-Excel Datei von SAP nach der Eingabe der Einzelantworten automatisch berechnet [74].

### 5.3.3 Resultate

In diesem Design Cycle haben die Testpersonen keine Probleme bei der Nutzung der Benutzeroberfläche gezeigt. Sie haben schnell und zielgerichtet gearbeitet, was ein Verständnis der Programmfunctionalitäten impliziert. Ein Nutzer hat jedoch gezeigt, dass das sofortige Löschen von allen Begrenzungsboxen durch das Klicken des entsprechenden Knopfes bei zu schneller Arbeit auftreten kann, da der Knopf direkt neben dem Hinzufügeknopf platziert ist. Jedoch wurden keine anderen Probleme oder technischen Fehler bei der Nutzung durch die fünf Testpersonen festgestellt.

Bei dem Fragebogen zur System-Gebrauchstauglichkeit haben die fünf nicht medizinischen Nutzer das Programm nach der dritten Nutzung meist positiv bewertet. Die Gesamtpunktzahl beträgt 71 was leicht über dem Durchschnitt liegt. Auf der Webseite des US Ministeriums für Gesundheit und Menschenservices wird eine Gesamtpunktzahl von 68 als Durchschnitt angegeben. Werte, die über dem Durchschnitt liegen, werden als gute Benutzbarkeit eingestuft und Werte, die darunter liegen, als schlechte Benutzbarkeit [76]. Die Standardabweichung der Gesamtpunktzahl beträgt dabei 37.3, da ein Nutzer das Programm deutlich schlechter bewertet hat als die anderen Nutzer. Bei der Wahl eines 95% Konfidenzniveaus liegt das Konfidenzintervall im Bereich [54.6, 87.4].

Die beiden neuen Testpersonen bewerten das Programm mit einem Gesamtwert von 68.8. Die Standardabweichung liegt dabei bei 12.4. Bei der Wahl des vorherigen Konfidenzniveaus wird ein Konfidenzintervall von [60.2, 77.3] erreicht. Damit liegt das Programm bei diesen beiden Testpersonen im Durchschnitt der Benutzbarkeit.

### 5.3.4 Gelerntes

Aufgrund der beobachteten Bearbeitungsgeschwindigkeit zeigt sich, dass das Programm bereits nach wenigen Benutzungen von den Testpersonen verstanden wird. Mithilfe des Fragebogens zeigt sich außerdem, dass die Probanden das Programm als gut nutzbar einschätzen und auch so verwenden würden, wenn sie die Aufgabe erhalten Bilder zu annotieren. Insgesamt lässt sich daraus ableiten, dass dieses Programm leicht zu verstehen, größtenteils fehlerfrei und nicht mit Funktionen überladen ist.

## 5.4 Design Cycle 4

### 5.4.1 Grundlagen und Implementierung

Die drei vorherigen Design Cycles befassen sich mit der Iteration über das Design der Annotationswebeite. In den nachfolgenden Design Cycles befasst sich die Arbeit mit dem tatsächlichen Training des ML Modells. In diesem und dem darauffolgenden Design Cycle wurden zwei verschiedene vortrainierte Modelle aus den von Tensorflow angebotenen Modelle verwendet. Dies ermöglicht es Unterschiede, die aus dem vorangegangen Training hervorgehen,

besser zu evaluieren und zudem ein Vergleichsmodell zu haben. Das erste Modell wurde auch beim ersten Prototyp, der während der Arbeit am Lehrstuhl der Wirtschaftsinformatik entstanden ist, verwendet. Dieses "faster\_rcnn\_inception\_v2\_cocoModell wurde, wie der Name es schon anzeigt, auf dem COCO Datensatz [35] trainiert und basiert auf der Faster R-CNN Architektur. Das Modell wurde mit der Pets-Konfiguration trainiert, welches im Gegensatz zur COCO-Konfiguration nur in der Evaluation der Sicherungspunkte des Trainings Unterschiede aufweist. Das zweite Modell, "faster\_rcnn\_inception\_resnet\_v2\_atrous\_oidv4", basiert auf dem Open Images Datensatz [34] und implementiert ebenfalls eine Faster R-CNN Architektur. In diesem Design Cycle wurden für das Training 364 Bilder der Medetec Datenbank verwendet [22], welche wie alle Bilder über den Dropbox Ordner dieser Arbeit heruntergeladen werden können [77]. Die Datenbank stellt Bilder in den drei zu untersuchenden Kategorien Dekubitus (163), venöse und arterielle Ulzera (121) und diabetischem Fußsyndrom (79) zur Verfügung. Aufgrund eines manuellen Fehlers beim Herunterladen der Bilder, wurde jedoch für das diabetische Fußsyndrom fälschlicherweise erneut das erste von zwei Paketen der venösen und arteriellen Ulzera genutzt. Auf diesen Fehler und seine Auswirkungen wird in Unterabschnitt 5.4.3 genauer eingegangen. Die Bilder wurden mit Hilfe von Philipp Reindl-Spanner [73] annotiert, welcher ebenfalls ein ML Modell für sein interdisziplinäres Projekt, einen webbasierten und ML gestützten Wunddokumentationsbogen für Pflegekräfte, an der TUM benötigt.

Es ist wichtig festzuhalten, dass aufgrund der SARS-CoV-2 Epidemie die Annotation nicht durch medizinisches Fachwissen validiert wurde, da dies ein großer Arbeitsaufwand ist und die während der Annotation stattfindende SARS-CoV-2 Epidemie noch im Prozess der Verschlechterung war. Zu diesem Zeitpunkt war eine mögliche Gefahr der Infektion bei einem Treffen mit medizinischem Personal vorhanden. Zudem war das medizinische Personal bereits bei einer hohen Arbeitsauslastung, weshalb eine weitere Arbeitsbelastung als nicht sinnvoll angesehen wurde. Für das Training wurden die Bilder in 80% Trainingsdaten und 20% Validierungsdaten geteilt. Für die anschließenden Performanztests wurden weitere 65 Bilder aus verschiedenen Quellen heruntergeladen, die alle entweder von medizinischem Personal stammen oder kontrolliert wurden, um die Korrektheit der Annotationen zu gewährleisten [24] [29] [78] [79] [80]. Nach der Annotation aller Bilder wurden die beiden ML Modelle trainiert, bis ihre Verlustfunktion keine signifikanten Absenkungen mehr aufzeigten (vgl. Abbildung 5.5, Abbildung 5.6). Die Verlustfunktion dient dabei als Indikator wie schlecht die Vorhersage des Modells gegenüber der Grundwahrheit bei jedem Trainingsschritt ist [81]. Beim COCO-basierten Modell trat dies nach 43119 Trainingsschritten ein, während beim Open Images-basierten Modell bereits nach 23228 Schritten keine großen Veränderungen, außer einem Ausbruch nach 22214 Schritten, zu sehen war. Im Anschluss an das Training wurde das Modell eingefroren, um es nach Außen nutzbar zu machen. Für die nachfolgenden Test wurde ein eigenes Evaluationsprogramm in Python geschrieben, das wichtige Messwerte wie die IoU, die Präzision, den Recall und F1-Wert berechnet [71].

Bei der Implementierung des ML Evaluationsprograms wurden dabei bewusst Entschei-

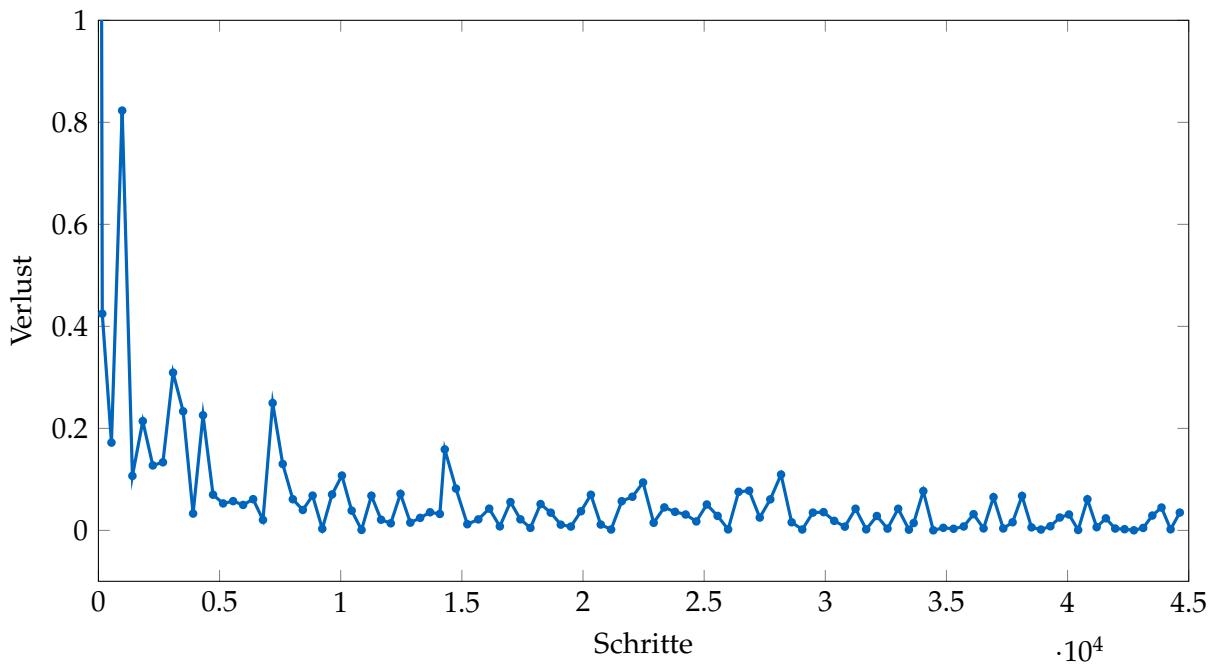


Abbildung 5.5: Verlustfunktion vom Trainingsprozess des auf COCO-basierten ML Modells für 43119 Schritte (Graph bei Verlust-Werten größer 1 ausgeblendet). (Design Cycle 4)

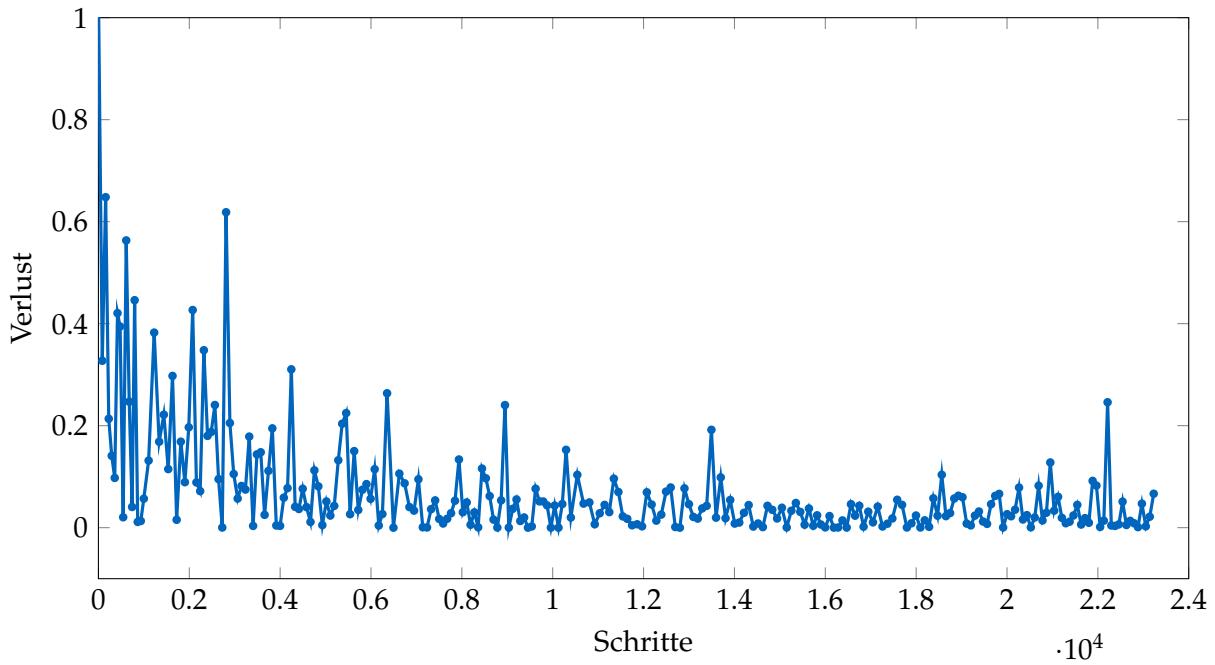


Abbildung 5.6: Verlustfunktion vom Trainingsprozess des auf Open Images-basierten ML Modells für 23228 Schritte (Graph bei Verlust-Werten größer 1 ausgeblendet). (Design Cycle 4)

dungen der Bewertung getroffen, die das Ergebnis gegenüber anderen Evaluationen sowohl positiv also auch negativ beeinflussen können. Denn je niedriger der Konfidenzschwellwert bei der Klassifikation durch das ML Modell gewählt wird, desto öfter treten Lokalisierungen mit geringer Konfidenz auf. Zudem müssen bei Bildern mit mehr als einer Wunde auch mehrere Lokalisierungen und Klassifizierungen stattfinden. Daraus entsteht die Frage wie die Vorhersagen und Grundwahrheiten miteinander verglichen werden. Eine Methode ist die Nutzung der ersten  $n$  Wundklassifikationen für  $n$  Wunden in der Grundwahrheit und Einordnung der verbleibenden Klassifizierungen als falsch positiv. Dies kann zu einem dramatischen Anstieg der falsch positiven Werte bei niedrigeren Konfidenzschwellwerten führen und die Gesamtperformanz stark verringern. Eine andere Methode ist die Priorisierung von Konfidenz über Lokalisierung. Dabei werden die  $n$  Klassifizierungen mit den höchsten Konfidenzen für die  $n$  Wunden in der Grundwahrheit herausgesucht und dann anhand der besten Lokalisierung zugeordnet. Dies führt dazu, dass vor allem unterschiedliche Objektregionen für die Bewertung genutzt werden. Denn zwei unterschiedliche Klassifizierungen, mit Konfidenzen von jeweils 99%, können nicht auf der exakt gleiche Objektregion zustande kommen, da dies dem grundsätzlichen Modell der regionsbasierten Klassifizierung widerspricht. Dies kann für mehrere Wunden sinnvoll sein und so immer Klassifizierungen mit hoher Konfidenz nutzen, aber auch zum Nachteil mit vielen kleinen erkannten Regionen werden. Denn wenn das RPN schlechte Regionen erkennt, und zum Beispiel viele kleine Wunden in einer Wunde erkennt, werden diese am Ende sehr wahrscheinlich unterhalb des Schwellwert der IoU landen. Eine dritte Möglichkeit der Auswahl stellt die Priorisierung von Lokalisierung über Konfidenz dar. Dabei wird für jede Wunde der Grundwahrheit die Vorhersage mit der höchsten IoU ermittelt und anschließend verglichen. Diese Methode erlaubt es Lokalisierungen von einer Wunde mit großer Wahrscheinlichkeit der Wunde in der Grundwahrheit zuzuordnen. Denn keine andere Vorhersage hat dadurch eine größere Überlappung mit der Grundwahrheit und damit eine bessere Chance den Schwellwert der IoU zu überschreiten. Sollten zwei Klassifizierungen der gleichen Region und damit gleicher IoU auftreten, wird anhand der Konfidenz entschieden welche Vorhersage verglichen wird. In dieser Arbeit wird zur Evaluation der Performanz der ML Modelle die Priorisierung der Lokalisierung über Konfidenz implementiert, da das übergeordnete Ziel ist, die Wunde nach der korrekten Lokalisierung richtig zu klassifizieren. Wenn die Lokalisierung bereits weit von der Grundwahrheit entfernt ist, kann dies auch zu einer falschen Klassifizierung führen, da nicht die gesamte Wunde erkannt wurde. Alle Vorhersagen die keine Wunde zugeordnet werden, werden in dieser Arbeit als falsch positiv markiert, da in der Praxis zwar eine Nachbearbeitung der Ergebnisse möglich ist, und bei höheren Konfidenzschwellwerten meist keine zusätzlichen Vorhersagen entstehen (vgl. Unterabschnitt 5.4.3), aber so nicht die direkte Performanz der ML Modelle widerspiegeln.

### **5.4.2 Tests**

Die Tests benötigen einige Messwerte, die in Abschnitt 3.2.1 beschrieben werden. Alle Messwerte basieren auf der Einordnung von Vorhersagen in die Wahrheitsmatrix (vgl. Abbildung 3.7). Aus diesem Grund erhielten alle Testbilder auch eine Annotation, um als Grundwahrheit zu dienen. Danach wurden alle Testbilder mit einer Abwandlung des Serverpro-

gramms direkt an das ML Modell geleitet, um eine möglichst effiziente Klassifikation zu ermöglichen. Die daraus entstandenen Vorhersagen, welche die Koordinaten der BB, den Wundtyp und die Konfidenz der Vorhersagen enthalten, wurden mit der Grundwahrheit an das Evaluationsprogramm weitergeleitet, um so die abschließenden Ergebnisse zu erhalten. Die Tests wurden für jedes Modell für zwei unterschiedliche IoU Schwellwerte durchgeführt. Dabei wurde der Standard von 0.5 sowie ein reduzierter Schwellwert von 0.3 herangezogen. Die Reduktion des Schwellwerts wurde als Ausgleich für die, nicht durch medizinisches Personal durchgeführte, Annotation der Trainings- und Testdaten angebracht. Für beide IoU wurden die Modelle für die zwei Konfidenzschwellwerte von 80% und 50% getestet. Dies erlaubt einen Einblick wie sich die Klassifizierung bei niedrigerer Konfidenz verändert und wie dies einen positiven Effekt auf die Messwerte, auch mit Blick auf den zuvor beschriebenen Fehler, hat. Neben diesen beiden manuell festgelegten Konfidenzschwellwerten wurden beide Modelle mit einem Schwellwert von 1% getestet um daraus die Präzision-Recall Kurve und Präzision-Recall-AUC zu berechnen. Die Präzision-Recall Kurve, statt einer ROC Kurve, ist ein gutes Mittel um die Performanz eines ML Modells, das sowohl klassifiziert als auch lokalisiert, zu messen.

#### 5.4.3 Resultate

Nachfolgend werden die Ergebnisse der Tests für die beiden ML Modelle des vierten Design Cycles beschrieben. Dabei werden für beide Modelle insgesamt vier Sets von Messwerten vorgestellt, sowie die Präzision-Recall Kurven und AUC. Denn eine Klassifizierung von rich-

Tabelle 5.2: Ergebnisse COCO basiertes ML Modell bei IoU Schwellwert = 0.3 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 4)

Klasse	Präzision	Recall	F1-Wert	Präzision	Recall	F1-Wert
DF	0	0	-	0	0	-
DK	0.5	0.667	0.571	0.500	0.611	0.550
VA	0.237	0.600	0.337	0.258	0.500	0.340
DK+VA	0.339	0.636	0.442	0.358	0.559	0.437
Gesamt	0.296	0.538	0.382	0.306	0.475	0.373

tig negativen Vorhersagen ist, ohne eine auf die Aufgabe spezifizierte Definition von richtig negativ, nicht realisierbar [82] [83]. Die in Tabelle 5.2 bis Tabelle 5.5 dargestellten Werte sind die Ergebnisse der Messungen für die beiden ML Modelle. Dabei steht 'DF' für diabetisches Fußsyndrom, 'DK' für Dekubitus, 'VA' für venöse und arterielle Ulzera, 'DK+VA' die Kombination der Einzelvorhersagen und 'Gesamt' für die Ergebnisse über alle Vorhersagen. Da alle Metriken einfach definiert sind, werden bei zusammengerechneten Werten die Summen über die Anzahl der Einordnungen in die Wahrheitsmatrix als Grundlage herangezogen, anstatt einen Durchschnitt zu berechnen. Bei allen Tests fällt das diabetische Fußsyndrom mit dem Wert Null für Präzision und Recall auf. Damit ist auch kein F1-Wert definiert, da ansonsten

Tabelle 5.3: Ergebnisse COCO basiertes ML Modell bei IoU Schwellwert = 0.5 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 4)

Klasse	Präzision	Recall	F1-Wert	Präzision	Recall	F1-Wert
DF	0	0	-	0	0	-
DK	0.458	0.647	0.537	0.455	0.588	0.513
VA	0.237	0.600	0.340	0.258	0.500	0.340
DK+VA	0.323	0.625	0.426	0.340	0.545	0.419
Gesamt	0.282	0.526	0.367	0.290	0.462	0.356

Tabelle 5.4: Ergebnisse Open Images basiertes ML Modell bei IoU Schwellwert = 0.3 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 4)

Klasse	Präzision	Recall	F1-Wert	Präzision	Recall	F1-Wert
DF	0	0	-	0	0	-
DK	0.778	0.636	0.700	0.778	0.636	0.700
VA	0.333	0.280	0.304	0.333	0.192	0.244
DK+VA	0.538	0.447	0.488	0.576	0.396	0.469
Gesamt	0.375	0.404	0.389	0.432	0.339	0.380

Tabelle 5.5: Ergebnisse Open Images basiertes ML Modell bei IoU Schwellwert = 0.5 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 4)

Klasse	Präzision	Recall	F1-Wert	Präzision	Recall	F1-Wert
DF	0	0	-	0	0	-
DK	0.611	0.579	0.595	0.611	0.579	0.629
VA	0.238	0.217	0.227	0.2	0.125	0.154
DK+VA	0.410	0.381	0.395	0.424	0.326	0.368
Gesamt	0.286	0.340	0.311	0.318	0.275	0.295

durch Null geteilt wird. In dieser Klasse wird kein einziges Testbild richtig klassifiziert. Dies ist nach dem Erkennen der falsch heruntergeladenen Bilder auch nicht überraschend, da die erkannten Wunden eigentlich zur Kategorie der venösen und arteriellen Wunden gehören. Die Auswirkungen des Fehler können sich zudem auch negativ auf die Performanz bei den venösen und arteriellen Ulzera auswirken. Denn Wundbilder die identisch zu den Bildern einer anderen Kategorie sind, werden als neue Klasse angegeben, was zum Versuch des ML Modells führen kann, die Unterschiede dieser zwei Gruppen in den identischen Bilder zu finden. Damit werden im Training falsche Schlüsse in das Modell trainiert, die dann eine Verschlechterung der Ergebnisse in der richtigen Klasse hervorrufen können. Wie stark dieser Einfluss ist, lässt sich nicht bestimmen, weil dafür das Modell genauestens untersucht und unter Umständen bis zu den Neuronengewichten analysiert werden muss. Dies bringt nur einen geringen Mehrwert, der allerdings viele Stunden Arbeit mit sich bringt. Eine einfachere Lösung ist das Trainieren eines neuen Modells mit den richtigen Bildern, was in Abschnitt 5.5 durchgeführt wurde.

Beim Vergleich der gleichen Modelle mit Unterschiedlichen Schwellwerten ist ein im vornherein vorhersagbares Ergebnis eingetroffen. In den meisten Fällen erreichen die Tests mit geringeren Schwellwerten bessere Ergebnisse als die Tests mit höheren Schwellwerten. Sowohl für die IoU als auch die Konfidenz bekommen bei geringeren Schwellwerten mehr Vorhersagen ein positives Testergebnis, was sich insgesamt positiv im Ergebnis niederschlägt. Für diese Regel gibt es in den Open Image basierten Tests bei der Präzision innerhalb des gleichen IoU Schwellwertes eine Besonderheit. Dort fällt der Messwert der Tests mit höherem Konfidenzschwellwert teils höher aus als bei einem niedrigeren Konfidenzschwellwert. Dies lässt dadurch erklären, dass Werte von falsch positiv zu falsch negativ umgewandelt werden, und damit das Ergebnis der Präzision verbessern, aber auch das Ergebnis des Recall verschlechtern. Mit dem Blick auf die einzelnen Klassen zeigt das Open Images basierte Modell eine deutliche Steigerung von ca. 55% bei der Präzision der Dekubituserkennung für einen IoU Schwellwert von 0.3 und Konfidenzschwellwert von 0.5 gegenüber dem äquivalenten Test beim COCO basierten Modell. Von 10 erkannten Dekubitus Bildern, enthalten fast 8 Bilder tatsächlich einen Dekubitus. Der Recall ist dafür aber niedriger, da dementsprechend einige weitere Bilder in die falsch negative Einordnung fallen und so die Gesamtzahl der als positiv erkannten von den tatsächlich positiven reduziert. Der F1-Wert ist dennoch deutlich höhere beim Open Images Modell als beim COCO Modell. Dies gilt in geringerem Ausmaß auch für die Tests mit höheren Schwellwerten. Mit Blick auf die venösen und arteriellen Ulzera zeigt sich, dass das COCO Modell bei diese Wunden sehr viele falsche Klassifikationen aufweist, weshalb die Präzision bei 0.237 liegt. Dennoch ist die Erkennungsrate der tatsächlich positiven Ulzera mit einem Recall von 0.6 bei fast zwei Dritteln. Das Open Images basierte Modell macht weniger falsch positive Erkennungen als das COCO basierte Modell, erkennt aber nur etwas mehr als ein Fünftel der Wunden aus der Grundwahrheit. Bei beiden Modellen liegt der F1-Wert deutlich unter 0.5 und mit großem Abstand hinter den Ergebnissen beim Dekubitus, welcher ein Maximum von 0.571 beziehungsweise 0.7 erreicht. Dementsprechend beeinflussen die beiden Werte auch die Ergebnisse bei der Zusammenfassung von Dekubitus

sowie venösen und arteriellen Ulzera. Die F1-Werte liegen je nach Wahl der Schwellwerte im Intervall [0.368, 0.488]. Mit Blick auf die Messwerte für die gesamten Modelle zeigen sich schlechte Ergebnisse. Der höchste F1-Wert, von 0.389 beziehungsweise 0.382, wird bei beiden Modellen für den niedrigeren IoU und Konfidenzschwellwert erreicht. Ein optimaler F1-Wert liegt bei 1.0 und ist damit deutlich vom erzielten Ergebnis entfernt. Besonders mit Blick auf das diabetische Fußsyndrom wird der F1-Wert im Vergleich zur Kombination des Dekubitus und der venösen und arteriellen Ulzera fast immer um den Wert 0.1 verringert.

Um beide Modelle abschließend miteinander zu vergleichen, sind in Abbildung 5.7 und Abbildung 5.8 die Präzision-Recall Kurven für einen IoU Schwellwert von 0.5 dargestellt. Der IoU Schwellwert 0.3 ist hier nicht mit einem Graphen dargestellt, da dieser einen Vergleich innerhalb der Arbeit zulässt, aber 0.5 oft in der Literatur verwendet wird. Diese Graphen erlauben es eine durchschnittliche Präzision zu berechnen. Dies geschieht mit der 11-Punkt Interpolation und der Gesamtpunktinterpolation, welcher ähnlich zur AUC bei einer ROC Kurve, die Fläche unterhalb des Graphen berechnet wird. Diese beiden Methoden wurden auch in dem PASCAL VOC Wettbewerb verwendet, und stellen damit eine gut Möglichkeit dar Modelle miteinander zu vergleichen [84]. Wie in Tabelle 5.6 zu sehen gibt es kleine

Tabelle 5.6: Durchschnittliche Präzision der Modelle sowie mAP, berechnet über die AUC der Präzision-Recall Kurven pro Klasse für die COCO (links) und Open Images (rechts) basierten Modelle. (Design Cycle 4)

Berechnungsart	DF	DK	VA	mAP	DF	DK	VA	mAP
11-Punkt	0.6%	44.4%	25.4%	23.5%	0%	42.8%	11.4%	18.1%
Gesamtpunkt	0.6%	45.6%	24.1%	23.4%	0%	53%	13.6%	22.2%

Unterschiede zwischen den beiden Messmethoden, da sie unterschiedlich stark approximieren. Bei 11-Punkt Interpolation werden nur die Präzisionswerte an 11 äquidistanten Punkten gemessen, addiert und dann durch 11 geteilt, um eine Prozentzahl zu erreichen. Bei der Gesamtpunktinterpolation wird von rechts beginnend die Fläche vom Recallwert mit der höchsten Präzision, bis zu einem linksliegenden Recallwert mit höherer Präzision berechnet. Danach wird der neue Recallwert genommen und bis zum nächsten Recallwert mit höherer Präzision die Fläche berechnet. Bei beiden Werten stellt 100% den Maximalwert dar und würden ein optimales ML Modell beschreiben. Bei den beiden vorgestellten Modellen erreicht ausschließlich das Open Images Modell eine durchschnittliche Präzision von über 50% für Dekubitus. Alle anderen Klassen und das COCO basierte Modell fallen unterhalb der 50% Schwelle und sind damit deutlich zu schlecht. Auch wenn es in diesem Gebiet kaum vergleichbare ML Modelle gibt, sind die mAPs mit deutlichen Abstand schlechter als ML Modelle in anderen Bereichen. Diese liegen aktuell bei 80% bis 95% (vgl. Abschnitt 3.3).

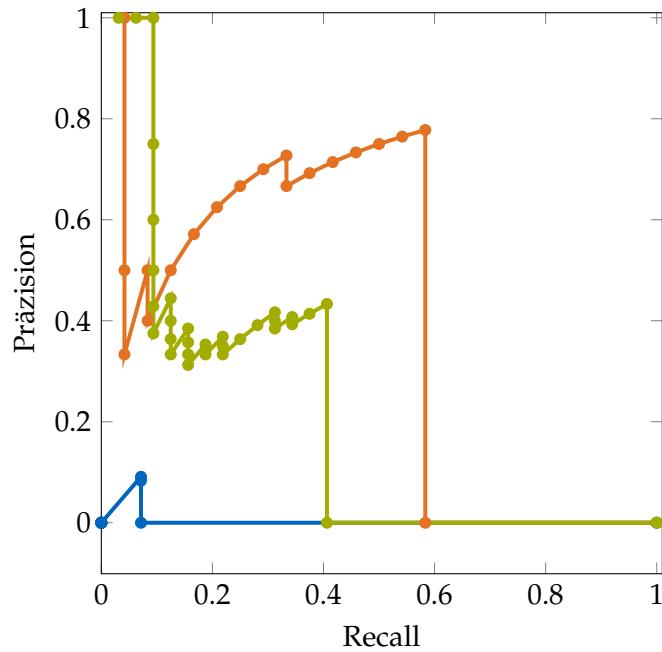


Abbildung 5.7: Präzision-Recall Kurven des COCO basierten Modells für diabetisches Fußsyndrom (blau), venöse und arterielle Ulzera (grün) und Dekubitus (orange) mit IoU = 0.5. (Design Cycle 4)

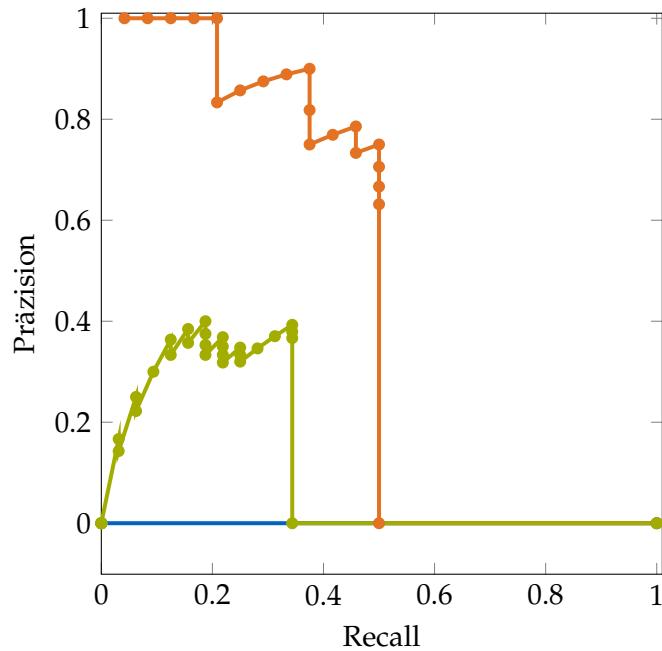


Abbildung 5.8: Präzision-Recall Kurven des Open Images basierten Modells für diabetisches Fußsyndrom (blau), venöse und arterielle Ulzera (grün) und Dekubitus (orange) mit IoU = 0.5. (Design Cycle 4)

#### 5.4.4 Gelerntes

Aus den vorangegangenen Ergebnissen lässt sich ableiten, dass die beiden ML Modelle deutlich unter dem aktuellen Stand der Technik liegen. Dies kann beim diabetischen Fußsyndrom auch auf die falschen Bilder zurückgeführt werden. Dennoch zeigen alle Messwerte, dass mehr Trainingsdaten benötigt werden, um mehr Diversität und damit bessere Lernerfolge erzielen zu können. Nach dem ersten Training ist erkennbar, dass sich die Dekubituswunden von den venösen und arteriellen Ulzera mit vergleichsweise hoher mAP unterscheiden lassen. Wie stark dieser Effekt bei zwischen dem diabetischen Fußsyndrom und den venösen und arteriellen Ulzera ist, kann zum derzeitigen Zeitpunkt nicht prognostiziert werden. Zudem ist es wichtig, verschiedene Messungen durchzuführen um verschiedene Szenarien der Nutzung beurteilen zu können und auch eine Präzision-Recall Kurve berechnen zu können. Damit lassen sich Parameter für den Einsatz leichter auswählen, als sie experimentell zu testen.

### 5.5 Design Cycle 5

#### 5.5.1 Grundlagen und Implementierung

Die im vorangegangenen Design Cycle erreichten Ergebnisse sind aufgrund des aufgetretenen Fehlers beim Herunterladen der Bilder nicht aussagekräftig genug um eine definitive Aussage über die ML Modelle zu machen. Daher wurden in diesem Design Cycle zwei weitere Modelle trainiert. Die beiden zugrundeliegenden Architekturen sind identisch zu den vorherigen. Zunächst wurden die korrekten Bilder für das diabetische Fußsyndrom heruntergeladen und annotiert. Um die dünne Datenlage des Datensatzes zu verbessern, wurde Bildaugmentation eingesetzt, um die Zahl der Bilder um den Faktor 13 zu erhöhen. Dies erlaubt es einen maximal ausgeglichenen Datensatz von jeweils 1500 Bildern pro Wundklasse zu erreichen und so mögliche Nachteile aus dem vorherigen Training auszuschließen. Für die Augmentation wurde die Python Bibliothek ImgAug [85] verwendet, welche viele verschiedene Bildeffekte enthält und zudem immer die BB entsprechend des Effekts anpasst. Das Ziel der Bildaugmentationen ist es mit der Veränderung von vorhandenen Bildern, neue Bilder zu erstellen [55] [86]. Diese müssen den grundlegenden Zusammenhang, welcher zwischen allen Bildern einer Klasse existiert und gelernt werden soll, weiterhin enthalten. Daher sind beliebte Methoden das Drehen, Spiegeln und leichte Anpassen von Farben in Bildern. In dieser Arbeit wurden die Bilder daher gedreht, gespiegelt, sowie den Bildern Helligkeit hinzugefügt, Rauschen überlagert, Kontrast angepasst und einzelne Farbkanäle leicht angepasst [71]. Im Anschluss wurden pro Klasse 1500 Bilder zufällig ausgesucht und zum Datensatz zusammengefasst. Dieser Datensatz wurde anschließend in 80% Trainingsdaten und 20% Validierungsdaten aufgeteilt.

Die beiden Graphen der Verlustfunktionen sind in Abbildung 5.9 und Abbildung 5.10 dargestellt. Wie im vorangegangenen Design Cycle musste das COCO basierte Modell deutlich mehr Trainingsschritte durchführen als das Open Images basierte ML Modell. Obwohl das Training des COCO Modells fast doppelt so viele Trainingsschritte absolviert hat, konvergiert

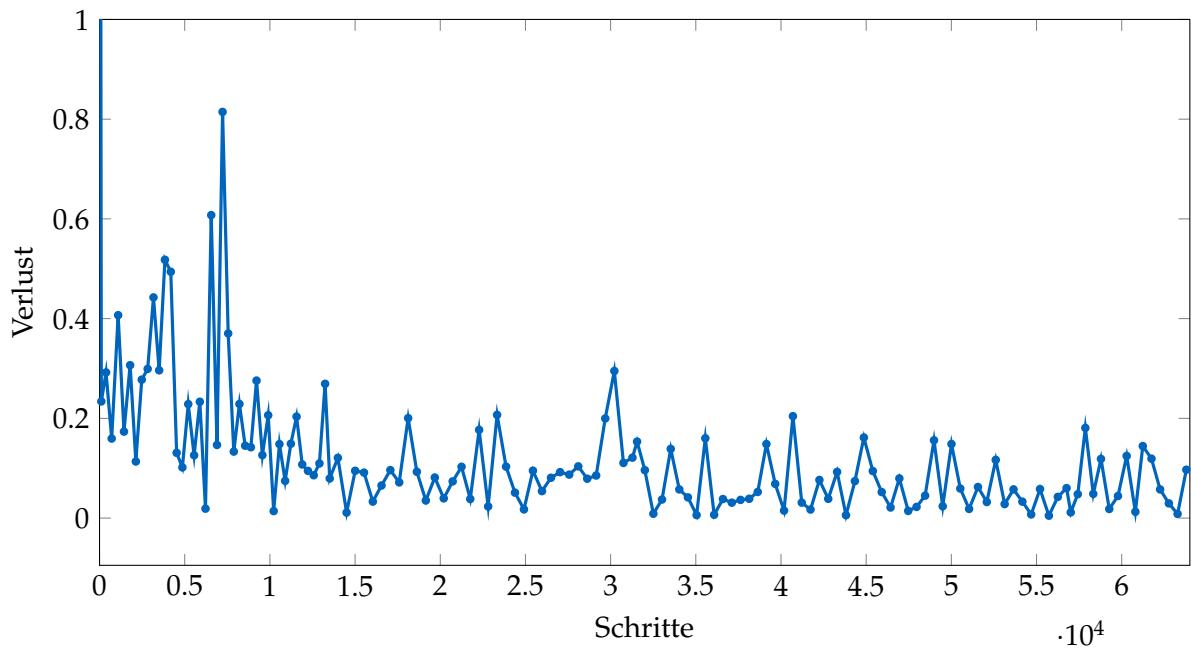


Abbildung 5.9: Verlustfunktion vom Trainingsprozess des auf COCO-basierten ML Modells für 63793 Schritte (Graph bei Verlust-Werten größer 1 ausgeblendet). (Design Cycle 5)

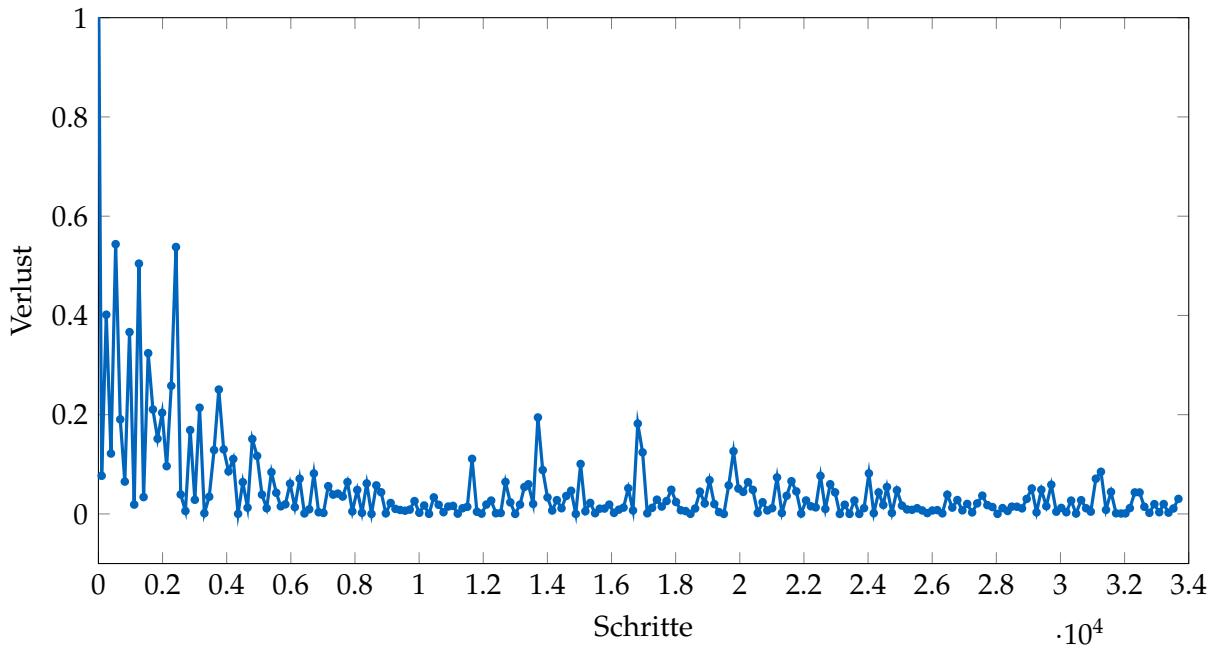


Abbildung 5.10: Verlustfunktion vom Trainingsprozess des auf Open Images-basierten ML Modells für 33678 Schritte (Graph bei Verlust-Werten größer 1 ausgeblendet). (Design Cycle 5)

das Open Images Modell zum Trainingsende deutlich stärker als das COCO Modell, welches weiterhin zwischen 0.1 und 0.01 divergiert. Hier könnte weiteres Training zu einer Konvergenz zu einem Wert unter 0.1 führen, was aber nicht eintreten muss. Daher wurde das Training nach mehrfacher Beobachtung der erneuten Divergenz nach 63793 Schritten beendet.

### 5.5.2 Tests

Nachdem das Training von beiden Modellen beendet wurde, wurden die Tests äquivalent zu den Tests aus Design Cycle 4 durchgeführt. Dieses Vorgehen erlaubt es einen guten Vergleich für alle Modelle zu erhalten. Zudem hat sich die Menge der Testbilder nicht vergrößert und eine Bildaugmentation von Testbildern könnte die Ergebnisse sowohl positiv als auch negativ beeinflussen.

### 5.5.3 Resultate

Die Ergebnisse des fünften Design Cycles sind in Tabelle 5.7 bis Tabelle 5.10 entsprechend dem vorangegangenen Schema dargestellt. Überraschenderweise sind die Ergebnisse im Vergleich zum vorhergehenden Design Cycle in den meisten Fällen schlechter ausgefallen. Mit Blick auf die nun korrekten Bilder des diabetischen Fußsyndroms zeigt sich gar kein Verbesserung beim COCO basierten Modell. In beiden Tests ist die Präzision und der Recall Null. Daraus lässt sich folgern, dass dieser Wundtyp, mit den bereitgestellten Bildern, für das COCO basierte Modell sehr schwer von den anderen Typen zu unterscheiden ist. Mit dem Blick auf die anderen Wundtypen, ist beim Dekubitus für beide IoU und Konfidenzschwellwerte jeweils ein Rückgang von 0.1 oder mehr im F1-Wert zu verzeichnen. Besonders der Recallwert zeigt deutliche Absenkungen zu den vorangegangenen Ergebnissen. Auch innerhalb eines IoU Schwellwerts verringert sich der Recall um mehr als 0.1 Punkte bei der Anhebung des Konfidenzschwellwerts. In den vorherigen Ergebnisse ist für den Dekubitus eine leichte Absenkung von wenigen hundertstel Punkten innerhalb eines IoU Schwellwerts zu verzeichnen. Die Präzision des Dekubitus sinkt ebenfalls zu den vorherigen Ergebnissen, aber nicht so signifikant wie der Recall oder F1-Wert. Im Gegensatz zum Dekubitus, steigt bei den venösen und arteriellen Ulzera die Präzision mit fast 0.1 Punkten für den geringeren Konfidenzschwellwert und etwas schwächer für den Höheren. Die bessere Präzision wird durch die stärkste Reduktion des Recallwerts beim COCO basierten Modell begleitet. Diese stellt mit 0.339 und 0.292 Punkten eine Verschlechterung um 56.5% beziehungsweise 58.4% im Vergleich zum Design Cycle 4 dar. Durch die Verbesserung der Präzision verschlechtert sich der F1-Wert zu den vorherigen Ergebnissen nur um 0.043 und 0.084 für die beiden unterschiedlichen Konfidenzschwellwerte. Mit dem Blick auf die aggregierten Ergebnisse zeigt sich bei den venösen und arteriellen Wunden ein Rückgang im F1-Wert von 0.06 Punkten für die beiden Messwerte mit geringerem Konfidenzschwellwert und ein Rückgang von fast 0.1 Punkten bei den Messwerten mit 80% als Konfidenzschwellwert. Wenn die Ergebnisse des diabetischen Fußsyndroms eingerechnet werden, so werden F1-Werte im Bereich 0.269 bis 0.330 erreicht. Über alle F1-Werte hinweg entsteht so ein durchschnittlicher Rückgang von 18.74%. Diese Ergebnisse sind mit Blick auf die deutlich höhere Datenmenge und des

Tabelle 5.7: Ergebnisse COCO basiertes ML Modell bei IoU Schwellwert = 0.3 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 5)

Klasse	Präzision	Recall	F1-Wert	Präzision	Recall	F1-Wert
DF	0	0	-	0	0	-
DK	0.435	0.526	0.476	0.471	0.400	0.432
VA	0.333	0.261	0.297	0.333	0.208	0.256
DK+VA	0.390	0.381	0.386	0.406	0.295	0.342
Gesamt	0.356	0.308	0.330	0.361	0.241	0.289

Tabelle 5.8: Ergebnisse COCO basiertes ML Modell bei IoU Schwellwert = 0.5 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 5)

Klasse	Präzision	Recall	F1-Wert	Präzision	Recall	F1-Wert
DF	0	0	-	0	0	-
DK	0.391	0.500	0.439	0.412	0.368	0.389
VA	0.333	0.261	0.297	0.333	0.208	0.256
DK+VA	0.366	0.366	0.366	0.375	0.279	0.320
Gesamt	0.333	0.294	0.313	0.333	0.226	0.269

Tabelle 5.9: Ergebnisse Open Images basiertes ML Modell bei IoU Schwellwert = 0.3 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 5)

Klasse	Präzision	Recall	F1-Wert	Präzision	Recall	F1-Wert
DF	0.059	0.333	0.1	0	0	-
DK	0.55	0.579	0.564	0.667	0.476	0.556
VA	0.25	0.3	0.272	0.2	0.143	0.167
DK+VA	0.386	0.436	0.410	0.433	0.310	0.361
Gesamt	0.295	0.429	0.350	0.371	0.245	0.295

Tabelle 5.10: Ergebnisse Open Images basiertes ML Modell bei IoU Schwellwert = 0.5 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 5)

Klasse	Präzision	Recall	F1-Wert	Präzision	Recall	F1-Wert
DF	0.059	0.333	0.1	0	0	-
DK	0.55	0.579	0.564	0.667	0.476	0.556
VA	0.167	0.222	0.190	0.133	0.100	0.114
DK+VA	0.341	0.405	0.370	0.4	0.293	0.338
Gesamt	0.262	0.400	0.317	0.343	0.231	0.276

deutlich längere Trainings als im Vergleich zum vierten Design Cycle überraschend.

Im Gegensatz zum COCO basierten Modell, kann das Open Images basierte Modell einige Wunden des diabetischen Fußsyndroms mit einer Konfidenz von weniger als 50% erkennen. Denn für beide IoU Schwellwerte ist bei geringerer Konfidenz eine Präzision und ein Recallwert vorhanden. Mit einer Präzision von 0.059 werden bei einem Recallwert von 0.333 viele falsch positive Vorhersagen gemacht. Denn ein Drittel aller tatsächlichen diabetischen Ulzera werden erkannt, aber nur jede zwanzigste erkannte Wunde fällt auch in diesen Wundtyp. Mit einem F1-Wert von 0.1 liegt die Performanz für diesen Wundtyp dennoch weit von einem guten Ergebnis entfernt. Ebenso wie beim COCO basierten Modell, fallen bei diesem Modell die Ergebnisse für den Dekubitus in allen Tests außer dem Test, welcher mit einem IoU Schwellwert von 0.5 und einem Konfidenzschwellwert von 80% durchgeführt wurde. Dort steigt die Präzision über denselben Wert im vorangegangenen Design Cycle mit 0.056 Punkten. Besonders bei den Test mit niedrigerem IoU Schwellwert ist ein größerer Unterschied der Präzision für den Dekubitus zu sehen. Der Unterschied beträgt bei geringerer Konfidenz 0.228 Punkte und bei höherer Konfidenz 0.109 Punkte. Beim Recall ist ein ähnliches Ergebnis wie bei der Präzision zu beobachten, was auch in einem schlechteren F1-Wert resultiert. Dieser verschlechtert sich um 0.136 und 0.144 Punkte im leichteren Test und einige Punkte weniger im Test mit höherem IoU Schwellwert. Bei den venösen und arteriellen Ulzera zeigen sich weitere Verschlechterungen der vorherigen schlechten Ergebnisse bei der Präzision. In den Test mit IoU Schwellwert von 0.5 fallen die Präzisionen unter 0.2, welches das schlechteste Ergebnis aus dem Design Cycle darstellt. Im Gegensatz zur Präzision, erreicht das Open Images basierte Modell leichte Anstiege bei dem Recallwert, wenn der Konfidenzschwellwert bei 50% gesetzt wird. Durch die Anhebung des Schwellwerts sinkt der Recallwert in beiden Tests mehr als 0.1 Punkte unterhalb der Ergebnisse aus dem vierten Design Cycle. Bei der Aggregation der Ergebnisse der Dekubitus und venösen und arteriellen Wunden fallen die meisten Ergebnisse schlechter als die Werte aus dem vorherigen Design Cycle aus. Bei den Tests mit geringeren IoU Schwellwert ist der Rückgang der Werte besonders bei der Präzision erkennbar. Diese fällt bei beiden Konfidenzschwellwerten mehr als 0.1 Punkte. Bei den beiden Recallwerte fällt dieser Unterschied kleiner aus und erreicht eine maximale Differenz

von 0.086 Punkten bei höherer Konfidenz. Entsprechend ist der Unterschied der F1-Werte geringer. Bei einem höheren IoU Schwellwert erreichen die Kombination von Dekubitus und venösen und arteriellen Ulzera einen höheren Recallwert bei dem Konfidenzschwellwert von 50%. Alle anderen Werte fallen um einige hundertstel Punkte von den Ergebnissen des vorherigen Design Cycles ausgehend. Mit dem Blick auf die Gesamtergebnisse des Open Images basierten Modells zeigt sich, dass die Messwerte durch die zusätzlichen Werte bei der Erkennung des diabetischen Fußsyndroms nahe an die Ergebnisse des vorherigen Design Cycle herankommen, wenn ein geringerer Konfidenzschwellwert gewählt wird. Bei den beiden Recallwerten der niedrigeren Konfidenz ist die Performanz des Modells besser als zuvor. Durch das Heraufsetzen des Konfidenzschwellwerts verlieren diese Werte aber auch einige Punkte, da dann auch keine Wunden des diabetischen Fußsyndroms erkannt werden und damit die schlechteren Ergebnisse beim Dekubitus und den venösen und arteriellen Ulzera nicht mehr ausgleichen. Beim Blick auf den Performanzverlust über alle F1-Werte hinweg, lässt sich eine Verringerung der Performanz von 9.96% feststellen, auch wenn das Modell aus dem vierten Design Cycle für einen IoU Schwellwert von 0.5 und einem Konfidenzschwellwert von 50% einen leichten Rückstand von 0.006 Punkten hat. Bei dem Open Images Modell lässt sich dennoch eine deutlich stärkere Konfidenz bei den Vorhersagen, sowie eine sehr stabile Lokalisierung von Objekten festhalten. Denn beim Dekubitus gibt es nur minimale Verschlechterungen bei der Erhöhung des Konfidenzschwellwerts, während es beim COCO basierten Modell auch bei der Erhöhung des IoU Schwellwerts eine Verschlechterung gibt.

Wie im vorherigen Design Cycle sind in Abbildung 5.11 und Abbildung 5.12 die Präzision-Recall Graphen dargestellt. Die mAP Werte für die beiden Graphen sind in Tabelle 5.11 dargestellt. Beim Vergleich der Werte mit den Ergebnissen des vorherigen Design Cycles zeigt sich in der mAP bei beiden Modellen ein Rückgang. Das COCO basierte Modell verschlechtert sich von 23.5%, beziehungsweise 23.4% auf 21.1% und 20.15%. Das Open Images basierte Modell weiß eine ähnliche Verschlechterung von 18.1%, beziehungsweise 22.2% auf 16.98% und 21.88%. Die zuvor ungenügenden Ergebnisse sind damit nicht verbessert worden, sondern liegen damit noch stärker von der aktuellen Spitze im Bereich ML entfernt.

Tabelle 5.11: Durchschnittliche Präzision der Modelle sowie mAP, berechnet über die AUC der Präzision-Recall Kurven pro Klasse für die COCO (links) und Open Images (rechts) basierten Modelle. (Design Cycle 5)

Berechnungsart	DF	DK	VA	mAP		DF	DK	VA	mAP
11-Punkt	0%	41.78%	21.51%	21.10%		0.54%	44.41%	5.98%	16.98%
Gesamtpunkt	0%	40.42%	20.02%	20.15%		1.79%	55.96%	7.89%	21.88%

#### 5.5.4 Gelerntes

Die Ergebnisse des fünften Design Cycles sind der Erwartung entgegen schlechter ausgefallen als im vierten Design Cycle. Besonders mit Blick auf die Funktionsweise von CNNs können

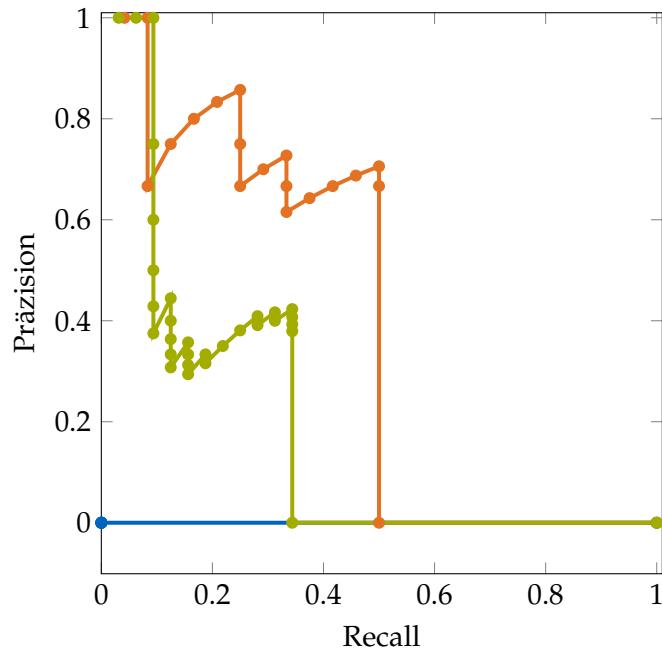


Abbildung 5.11: Präzision-Recall Kurven des COCO basierten Modells für diabetisches Fußsyndrom (blau), venöse und arterielle Ulzera (grün) und Dekubitus (orange) mit IoU = 0.5. (Design Cycle 5)

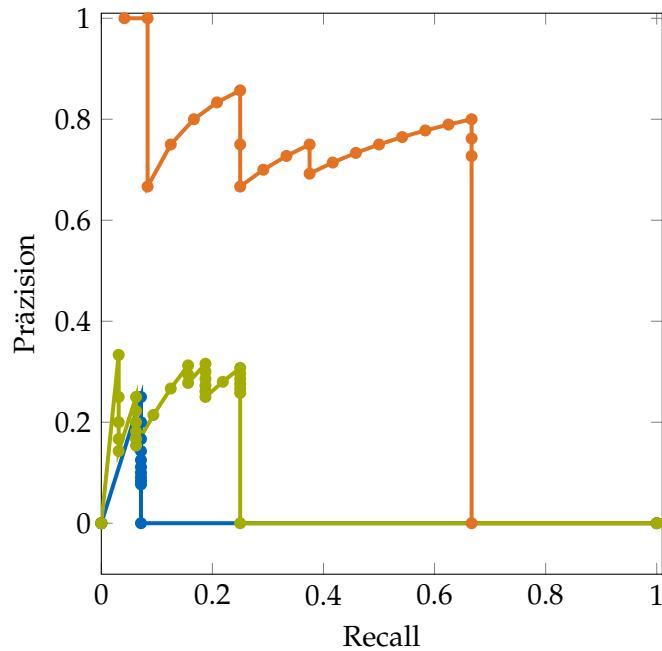


Abbildung 5.12: Präzision-Recall Kurven des Open Images basierten Modells für diabetisches Fußsyndrom (blau), venöse und arterielle Ulzera (grün) und Dekubitus (orange) mit IoU = 0.5. (Design Cycle 5)

bei mehr Daten, mit gleichem Qualitätsgrad, normalerweise bessere Ergebnisse erzielt werden. Die schlechten Ergebnisse des COCO basierten Modells lassen sich zu einem gewissen Grad auf die Divergenz der Verlustfunktion während des Trainingsprozesses zurückführen, welcher zu keinem Zeitpunkt eine starke Konvergenz wie beim Open Images Modell zeigte. Dennoch ist auch beim Open Images basierten Modell eine Verschlechterung der Ergebnisse zu sehen, obwohl die Verlustfunktion stark konvergiert ist. Es ist möglich, dass bei den beiden Modellen Überanpassung eingetreten ist, weil zu lange trainiert wurde. Nach Bishop [87] kann bei Neuronalen Netzwerken Überanpassung erkannt werden, wenn die Verlustfunktion zunächst sinkt und dann nach einem längeren Minimum wieder ansteigt. Dies ist beim COCO basierten Modell möglich, da die Verlustfunktion nicht dauerhaft auf ein Minimum absinkt. Allerdings gab es auch keinen Zeitraum, zu dem das Modell für einen längeren Trainingszeitraum stabil war. Im Gegensatz zum COCO basierten Modell ist das Open Images basierte Modell sehr schnell abgesunken und bei bereits 10000 Schritten stabil auf einem Minimum geblieben. Danach hat die Verlustfunktion begonnen wieder anzusteigen und über einen Wert von 0.2 zu steigen. Dies könnte sehr stark auf Überanpassung hinweisen, da das Modell bereits für einige Zeit stabil war und danach wieder größere Fehler gemacht hat.

Dieser Design Cycle zeigt ebenfalls, dass die Bildaugmentation in diesem Fall nicht so gut funktioniert, wie es in anderen wissenschaftlichen Arbeiten angegeben wird (vgl. Abschnitt 3.3). Dies kann mehrere Gründe haben. Zum einen können die Änderungen nicht stark genug sein, so dass das Modell keine Mehrinformationen aus den neuen Bildern herauszieht. Aus diesem Grund muss überprüft werden, ob stärkere Bildeffekte einen Vorteil bieten, oder zu viel Rauschen in das Modell hineinbringen. Ein anderes potenzielles Problem ist, dass die grundlegenden Bilder kaum bessere Ergebnisse zulassen. Denn wenn ein Modell bei den grundlegenden Bildern keine gemeinsamen Merkmale zwischen den Bildern finden kann, muss der Verbesserungseffekt durch Bildaugmentation nicht sehr groß sein.

Insgesamt sind die Ergebnisse des Design Cycle als Erkennungssystem unbrauchbar, da ein System, das in den meisten Fällen keine oder sogar falsche Ergebnisse liefert, nicht im Alltag einsetzbar ist. Dies gilt für alle Bereiche, aber mit besonderem Nachdruck für die Medizin, in welcher falsche Entscheidungen zu einer falschen Behandlung und daraus resultierenden Problemen für die Patienten führen kann.

# **6 Implikationen**

Die Ergebnisse aus den vorangegangenen Design Cycles zeigen einige Möglichkeiten auf, die nachfolgend genauer analysiert werden. Zunächst werden theoretisch, wissenschaftliche Implikationen beschrieben und anschließen die Implikationen für die praktische Anwendung.

## **6.1 Theoretische Implikationen**

Die Ergebnisse des vierten und fünften Design Cycles haben besonders theoretische Implikationen für diese Arbeit. Durch die schlechten Ergebnisse lässt sich auf dem Gebiet des ML vor allem aufzeigen, dass generische, open source CNN Modellarchitekturen für das Gebiet der chronischen Wunden, welches mit wenigen Daten und hohen Anforderungen aufwartet, ungeeignet sind. Die Nutzung aktueller Technologien und speziell auf das Problem abgestimmte CNN Modelle müssen zeigen, wie ein zuverlässiger Einsatz von ML in der medizinischen Bildverarbeitung aussehen kann. Mit dieser Arbeit wurde eine klare Einordnung von frei zugänglicher ML Technologie in den wissenschaftlichen Raum des ML vorgenommen. Es ist damit deutlich erkennbar, welche Technologien für zukünftige Technologien bereitstehen und welche nicht.

Mit Blick auf die Annotationsplattform zeigen sich spannende Vereinfachungskonzepte, mit denen medizinisches Personal für technische Hilfsmittel gewonnen werden können. Besonders kleine und schnell nutzbare Programme sind hier wichtig.

## **6.2 Praktische Implikationen**

Besonders mit Blick auf die Annotationsplattform hat diese Arbeit wichtige praktische Implikationen für die Pflege aber auch für die Wissenschaft. In der Pflege ist das Programm leicht einsetzbar, da es sehr effizient arbeitet und mit grundlegenden Bedienkonzepten eines Computers arbeitet. Dadurch werden alle Pflegekräfte einer Institution in die Lage versetzt das Programm lernen und dann in den Alltag integrieren zu können. Dadurch kann die große Menge an fachlichem Wissen, die durch medizinisches Personal bereitgestellt wird, für die Verbesserung des Alltags des medizinischen Personals genutzt werden. Durch solch einen selbstverstärkenden Prozess sind neue Datensätze, die zuvor nicht möglich waren, in die Nähe der medizinischen Institutionen und Wissenschaft gerückt. Denn auch die Wissenschaft kann durch neue Datensätze profitieren, indem sie mit bereits bestmöglich annotierten Daten in allen Bereichen neue Lösungen für aktuelle Probleme entwickeln kann. Diese können von weiteren ML Systemen für medizinisches Personal bis zur Entwicklung

## *6 Implikationen*

---

neuer Produkte und Heilverfahren für Patienten reichen. Auch außerhalb des medizinischen Bereiches gibt es viele Anwendungsmöglichkeiten, die auf annotierte Bilddaten zurückgreifen. Mit leichten Änderungen und beliebiger Erweiterbarkeit des Programms ist so ein Transfer in beliebige Bereiche der Wissenschaft möglich.

Die Einsatzmöglichkeiten des Wunderkennungssystems sind mit den aktuellen Ergebnissen nicht gegeben. Es müssen deutliche Verbesserungen in Richtung des aktuellen Stands der Wissenschaft im Bereich des ML passieren, um eine System in den medizinischen Bereich bringen zu können. Trotz der insgesamt schwachen Ergebnisse hat das System aber auch gezeigt, dass eine Klasse, nämlich der Dekubitus, eine gute Grundlage bietet für die Verbesserung des Systems.

## 7 Zukünftige Forschungsfragen

Um die Klassifizierung von chronischen Wunden in Zukunft mit ML zu realisieren, müssen einige wichtige Punkte realisiert werden. Zunächst muss die Performanz von ML gestützten Klassifizierungssystemen deutlich verbessert werden. Dafür werden Systeme mit mAP von über 90% benötigt, die auch in weiteren Messwerten optimale Ergebnisse liefern können. Des Weiteren muss festgestellt werden, wie viele weitere Wundtypen klassifiziert werden können. Denn neben den vier in dieser Arbeit vorgestellten Wundarten gibt es noch einige weitere, die entweder sehr selten vorkommen, oder in andere Kategorien von Wunden fallen, wie zum Beispiel Brandwunden. Des Weiteren stellen sich in Bezug auf das gemeinsame Auftreten von Wunden die Fragen wie damit umgegangen wird. Hinzu kommen auch die in Abschnitt 3.1 beschrieben Schweregraden von Wunden, welche in weiterer Forschung unterschieden werden könnte. Dabei steht die Implementierung neuer Objektklassen, die dann jede mögliche Kombination von Wunden und Schweregraden abdecken, der Nutzung von einer ML Klassifizierungspipeline gegenüber. In solch einer Pipeline wäre es denkbar pro Wundtyp ein ML System zu entwickeln, dass nur diese eine Wunde mit den verschiedenen Schweregraden erkennen kann und dann aus allen Vorhersagen die endgültige Vorhersage zu akkumulieren.

Eine weitere offene Forschungsfrage ist die Parallelisierung von ML Erkennungssystemen für eine Pipeline, aber auch für ein einzelnes Modell. Denn bei der Nutzung der Annotationswebseite wird nach jedem Hochladen das ML Modell aktiviert. Da dies aber nur einmal gleichzeitig pro Installation laufen kann, werden alle anderen Nutzer, die auch ein Bild hochladen in eine Warteschlange eingereiht, und müssen unter Umständen lange auf die Ergebnisse warten. Dafür wäre ein parallel ausführbares ML Modell sinnvoll, welches die Wartezeiten möglichst verhindern kann. Aber auch andere Systeme wie das Sammeln und Vorhersagen von Bildern bevor die Annotation stattfindet.

Abschließend steht die Einbindung des ML Systems in einen realen Arbeitsprozess im Krankenhaus aus. Diese muss auf der Basis eines anderen Systems, welches die Komponente mit aufnehmen kann, geschehen. Die anschließende Nutzungsbereitschaft hängt von der Umsetzung des Trägersystems ab, welches möglichst störungs- und irritationsfrei sein sollte. Eine mögliche Forschungsarbeit besteht darin, verschiedene Systeme aus vorangegangenen Arbeiten zusammenzusetzen und zu prüfen, auf welchen Systemen eine einfache Einbindung sichergestellt werden kann.

## 8 Abschließende Zusammenfassung

In dieser Bachelorarbeit wurde auf der Grundlage der Digitalisierung des Wundpflegeprozesses ein neues System zur automatischen Erkennung von Wundtypen entwickelt. Mit Blick auf andere Forschung in diesem Bereich, ist diese Arbeit eine von wenigen die sich mit der Klassifizierung von Wundtypen beschäftigt. Für das implementierte System wurde beispielhaft ein Prozess, der auf einem Prozess der Augmented Reality Pflege basiert, modelliert und anschließend das alleinstehende System in zwei Iterationen getestet. Ein automatisches Wunderkennungssystem hat die Fähigkeit Bewertungen und Dokumentationen von Wunden für die Pflegekräfte zu vereinfachen und zu vereinheitlichen. Die Ergebnisse der Implementierung erlauben aber keine direkte Anwendung in der praktischen Anwendung, sondern müssen zunächst in zukünftigen Arbeiten verbessert werden. Verbesserungen müssen in Zukunft durch speziell für die Anwendung angepasste ML Modelle geschaffen werden, da der aktuelle Stand der Technik im Bereich des Open Source ML in Form des Tensorflow Framework und aktuelle Open Source Modellarchitekturen, nicht die benötigten Ergebnisse geliefert hat.

Im Zuge der Arbeit gab es besonders bei der Datensammlung große Probleme, da kaum öffentlich zugängliche und dazu medizinisch verifizierte Wundbilder vorhanden sind. Um die fehlenden Daten für einen heterogenen ML Trainingsdatensatz in Zukunft ermöglichen zu können, wurde eine webbasierte Plattform zur Annotation von Wundbildern entwickelt. Diese wurde in drei Iterationen getestet, verbessert und erreichte im Anschluss ein gutes Ergebnis im System-Gebrauchstauglichkeit Test. Dieses System soll dabei helfen, mit wenig Aufwand für die Pflegekräfte einen großen Gewinn an medizinisch verifizierte Bilddaten für zukünftige ML Projekte zu generieren, indem es das Wissen von ausgebildetem Personal nutzt. Auch dafür wurde ein einfacher Prozessablauf modelliert. Abschließend bleibt noch der tatsächliche Einsatz des Systems in einer medizinischen Umgebung offen, welche nach dem Ende der zur Arbeit stattfindenden Pandemie wieder möglich sein wird.

Das Feld der Wundpflege bleibt weiterhin ein wichtiger Teil der Forschung im Bereich der Informationssysteme und wird sich in den nächsten Jahren durch Legislatur, Forschung und neue Technologien stark weiterentwickeln.

# Abbildungsverzeichnis

2.1	Darstellung der einzelnen Cycles der Design Science Methode in dieser Arbeit	4
3.1	Dekubitus mit einer partiellen Nekrose . . . . .	8
3.2	Ulkus infolge einer Erkrankung an diabetischem Fußsyndrom . . . . .	9
3.3	Ulkus infolge einer Veneninsuffizienz . . . . .	11
3.4	Leichter Ulkus aufgrund einer arteriellen Insuffizienz . . . . .	11
3.5	Beispielhafter Graph für ein überwachtes ML System mit Regressionsproblem.	13
3.6	Beispielhafter Graph für ein ternär-klassifizierendes überwachtes ML System.	13
3.7	Wahrheitsmatrix mit den möglichen Einordnungen von Klassenvorhersagen zur Grundwahrheit eines ML Modells. . . . .	15
3.8	Beispielhafte Berechnung der IoU. . . . .	16
3.9	Beispielhafte ROC Kurve (blau), dazugehörige AUC (grau) und Mittelkurve (orange) . . . . .	18
3.10	Darstellung eines 2-Ebenen Neuronalen Netzes mit vollständig verbundenen Ebenen. . . . .	19
3.11	Rectified Linear Unit Aktivierungsfunktion . . . . .	20
3.12	Darstellung einer überangepassten Funktion im Vergleich zu einer generalisierten Funktion . . . . .	21
4.1	Wundpflegeprozess mit Augmented Reality Brille nach Klinker et al. [66] . . .	31
4.2	Um ML gestützte Wundklassifizierung erweiterter Prozessablauf für die Schritte 'Wundgröße mit der HoloLens aufnehmen' bis 'Wunde dokumentieren' aus Abbildung 4.1. . . . .	32
4.3	Annotationsprozess von Wundbildern durch eine Pflegekraft. Orange markierter Schritt wird übersprungen, falls noch kein Wunden erkennendes ML Modell vorhanden ist. . . . .	33
5.1	Startseite der Klassifizierungswebseite. (Design Cycle 1) . . . . .	36
5.2	Eingabeoberfläche des Klassifizierungsprogramms, um Begrenzungsboxen und Labels zu bearbeiten. (Design Cycle 1) . . . . .	37
5.3	Eingabeoberfläche des Klassifizierungsprogramms, um Begrenzungsboxen und Labels zu bearbeiten. (Design Cycle 2) . . . . .	40
5.4	Hilfefenster mit Erklärungen der Funktionsweise des Programms. (Design Cycle 2) . . . . .	41
5.5	Verlustfunktion vom Trainingsprozess des auf COCO-basierten ML Modells für 43119 Schritte. (Design Cycle 4) . . . . .	46

## *Abbildungsverzeichnis*

---

5.6	Verlustfunktion vom Trainingsprozess des auf Open Images-basierten ML Modells für 23228 Schritte. (Design Cycle 4) . . . . .	46
5.7	Präzision-Recall Kurven des COCO basierten Modells. (Design Cycle 4) . . . . .	52
5.8	Präzision-Recall Kurven des Open Images basierten Modells. (Design Cycle 4) . . . . .	52
5.9	Verlustfunktion vom Trainingsprozess des auf COCO-basierten ML Modells für 63793 Schritte. (Design Cycle 5) . . . . .	54
5.10	Verlustfunktion vom Trainingsprozess des auf Open Images-basierten ML Modells für 33678 Schritte. (Design Cycle 5) . . . . .	54
5.11	Präzision-Recall Kurven des COCO basierten Modells. (Design Cycle 5) . . . . .	59
5.12	Präzision-Recall Kurven des Open Images basierten Modells. (Design Cycle 5) . . . . .	59

# Tabellenverzeichnis

3.1 Tabelle mit Erklärungen für die vier üblichen Einordnungen von Dekubitus. . . . .	8
3.2 Tabelle mit Einordnung der drei Arten von diabetischen Fußsyndrom. . . . .	9
3.3 Tabelle mit Erklärungen für die sieben üblichen Einordnungen von venösen Ulzera. . . . .	10
3.4 Tabelle mit Erklärungen für die vier üblichen Einordnungen von arterieller Insuffizienz. . . . .	10
4.1 Tabelle mit wichtigen Beobachtungen für die Wundbeurteilung. . . . .	29
5.1 Tabelle mit allen Funktionen des Klassifizierungsprogramms. (Design Cycle 1)	37
5.2 Ergebnisse COCO basiertes ML Modell bei IoU-Schwellwert=0.3 & Konfidenzschwellwert=50% (links) und Konfidenzschwellwert=80% (rechts). (Design Cycle 4) . . . . .	48
5.3 Ergebnisse COCO basiertes ML Modell bei IoU-Schwellwert=0.5 & Konfidenzschwellwert=50% (links) und Konfidenzschwellwert=80% (rechts). (Design Cycle 4) . . . . .	49
5.4 Ergebnisse Open Images basiertes ML Modell bei IoU-Schwellwert=0.3 & Konfidenzschwellwert=50% (links) und Konfidenzschwellwert=80% (rechts). (Design Cycle 4) . . . . .	49
5.5 Ergebnisse Open Images basiertes ML Modell bei IoU-Schwellwert=0.5 & Konfidenzschwellwert=50% (links) und Konfidenzschwellwert=80% (rechts). (Design Cycle 4) . . . . .	49
5.6 Durchschnittliche Präzision der Modelle sowie mAP, berechnet über die AUC der Präzision-Recall Kurven pro Klasse für die COCO (links) und Open Images (rechts) basierten Modelle. (Design Cycle 4) . . . . .	51
5.7 Ergebnisse COCO basiertes ML Modell bei IoU Schwellwert = 0.3 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 5) . . . . .	56
5.8 Ergebnisse COCO basiertes ML Modell bei IoU Schwellwert = 0.5 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 5) . . . . .	56
5.9 Ergebnisse Open Images basiertes ML Modell bei IoU Schwellwert = 0.3 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 5) . . . . .	56

---

*Tabellenverzeichnis*

---

5.10 Ergebnisse Open Images basiertes ML Modell bei IoU Schwellwert = 0.5 & Konfidenzschwellwert = 50% (links) und Konfidenzschwellwert = 80% (rechts). (Design Cycle 5) . . . . .	57
5.11 Durchschnittliche Präzision der Modelle sowie mAP, berechnet über die AUC der Präzision-Recall Kurven pro Klasse für die COCO (links) und Open Images (rechts) basierten Modelle. (Design Cycle 5) . . . . .	58

# Glossar

**Begrenzungsbox** Eine rechteckige Umrandung die durch 2 Koordinatenpunkte in 2D beschrieben wird und ein assoziiertes Label hat. Markiert für Überwachtes Maschinelles Lernen den Bereich in dem das zu lernende Feature ist.. 35

**Ischämie** Eine Unterversorgung von Gewebe, welche bei längerem Anhalten zu dauerhaften Gewebeschäden, wie zum Beispiel Nekrose, führen kann. In Verbindung mit Druckeinwirkung kann eine Dekubituswunde entstehen.. 7

**Open Source** Programmcode der öffentlich einsehbar, herunterladbar und veränderbar auf einem öffentlich zugänglichen Server liegt.. 2

**Responsive Design** Ein HTML und CSS Design Format, in welchem mithilfe von relativen UI Layouts auf verschiedene Größen von Bildschirmen und Fenstern reagiert wird. Dadurch ist eine Webseite auf allen unterstützen Größen immer vollständig nutzbar und alle Elemente sichtbar.. 39, 41

# Akronyme

**AUC** Fläche unter der Kurve. 17, 18, 48, 51, 58, 65, 67, 68

**BB** Begrenzungsbox. 12, 16, 33, 34, 37, 48, 53

**CNN** Convolutional Neural Network. 18, 22, 23, 58, 61

**FPR** falsch positiv Rate. 17, 18

**IoU** Überschneidung über Vereinigung. 15, 16, 25, 45, 47–52, 55–59, 65, 67, 68

**mAP** gemittelte durchschnittliche Präzision. 18, 51, 53, 58, 63, 67, 68

**ML** Maschinelles Lernen. 2–5, 12–15, 17, 24, 26, 27, 29, 30, 32–36, 43–51, 53, 54, 56–58, 61–68

**ROC** Receiver operating characteristic. 17, 18, 48, 51, 65

**RPN** Regionsvorschlag-Netzwerk. 23, 24, 47

**RPR** richtig positiv Rate. 17, 18

**SVM** Stützvektormaschine. 23, 25

# Literatur

- [1] I. Köster und I. Schubert, „Epidemiologie und Versorgung von Patienten mit chronischen Wunden- Eine Analyse auf der Basis der Versichertenstichprobe AOK Hessen/KV Hessen - Zusammenfassung der Ergebnisse“, Köln, Techn. Ber., 2016. Adresse: <https://www.bvmed.de/download/pmv-zusammenfassung-der-ergebnisse>.
- [2] BVMed, „Wirtschaftlichkeit und Gesundheitspolitik: Einsatz von hydroaktiven Wundaflagen“, S. 11–13, 2015.
- [3] A. Büscher, P. Blumenberg, M. Krebs, M. Moers, A. Möller, D. Schiemann und H. Stehling, „Expertenstandard Pflege von Menschen mit chronischen Wunden“, Deutschen Netzwerk für Qualitätsentwicklung in der Pflege, Techn. Ber., 2015. Adresse: [https://www.dnqp.de/fileadmin/HSOS/Homepages/DNQP/Dateien/Expertenstandards/Pflege\\_von\\_Menschen\\_mit\\_chronischen\\_Wunden/ChronWu\\_Akt\\_Auszug.pdf](https://www.dnqp.de/fileadmin/HSOS/Homepages/DNQP/Dateien/Expertenstandards/Pflege_von_Menschen_mit_chronischen_Wunden/ChronWu_Akt_Auszug.pdf).
- [4] Institut für Gesundheits-System-Forschung, „Versorgung von Patienten mit chronischen Wunden in Deutschland - Ergebnisse einer Umfrage mit niedergelassenen Ärzten“, Techn. Ber., 2006. Adresse: <https://www.bvmed.de/de/bvmed/publikationen/broschueren-moderne-wundversorgung>.
- [5] B. für Gesundheit, *GKV-Versorgungsstrukturgesetz*, 2015. Adresse: *GKV-Versorgungsstrukturgesetz*.
- [6] H. Korzilius, „Ärztemangel: Den Nachwuchs aufs Land locken“, *Dtsch Arztebl International*, Jg. 112, Nr. 38, A–1504–, 2015. Adresse: <https://www.aerzteblatt.de/int/article.asp?id=172060>.
- [7] L. Malin, A. Jansen, S. Seyda und R. Flake, „Fachkräftesicherung in Deutschland – diese Potenziale gibt es noch“, KOFA, Techn. Ber., 2019.
- [8] B. f. B. und Forschung, *Digitalisierung in der Medizin*. Adresse: <https://www.bmbf.de/de/digitalisierung-in-der-medizin-2897.html> (besucht am 26.06.2020).
- [9] PWC, „How artificial intelligence may improve quality and efficiency, whilst reducing healthcare costs in Europe“, *Sherlock in Health: How artificial intelligence may improve quality and efficiency , whilst reducing healthcare costs in Europe*, Nr. June, S. 24, 2017.
- [10] Google, *TensorFlow*, 2020. Adresse: <https://www.tensorflow.org/> (besucht am 04.01.2020).
- [11] A. R. Hevner, „A three cycle view of design science research“, *Scandinavian journal of information systems*, Jg. 19, Nr. 2, S. 4, 2007.

- [12] C. Sonnenberg und J. vom Brocke, „Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research“, in *Design Science Research in Information Systems. Advances in Theory and Practice*, K. Peffers, M. Rothenerberger und B. Kuechler, Hrsg., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, S. 381–397, ISBN: 978-3-642-29863-9.
- [13] C. Lewis und J. Rieman, *Task-Centered User Interface Design A Practical Introduction*. 1993. Adresse: <https://www.hcibib.org/tcuid/>.
- [14] J. Nielsen, *Why you only need to test with 5 users*, 2000. Adresse: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> (besucht am 31.01.2020).
- [15] K. Zulkowski, „Wound terms and definitions“, *WCET Journal*, Jg. 35, Nr. 1, S. 1–137, 2015. DOI: 10.4324/9780429268144-1.
- [16] P. Runge-Werner, J. Hoffmann und J. Dissemond, *Wundfibel*. Essen, 2015, S. 1–60. Adresse: [https://www.uk-essen.de/fileadmin/Hautklinik/Fuer\\_AErzte/Wundfibel\\_01.pdf](https://www.uk-essen.de/fileadmin/Hautklinik/Fuer_AErzte/Wundfibel_01.pdf).
- [17] C. Settelen, D. Stebler-Schärz und B. Egger, *Leitlinie Wundmanagement*. Universitätsspital Basel, 2008. Adresse: [https://www.unispital-basel.ch/fileadmin/unispitalbase\\_lch/Ressorts/Entw\\_Gesundheitsberufe/Abteilungen/Leitlinie\\_Wundmanagement/LLgesamtdokument.pdf](https://www.unispital-basel.ch/fileadmin/unispitalbase_lch/Ressorts/Entw_Gesundheitsberufe/Abteilungen/Leitlinie_Wundmanagement/LLgesamtdokument.pdf).
- [18] S. Gupta, C. Andersen, J. Black, J. De Leon, C. Fife, J. Ii, J. Niegzoda, R. Snyder, B. Sumpio, W. Tettelbach, T. Treadwell, D. Weir und R. Silverman, „Management of Chronic Wounds: Diagnosis, Preparation, Treatment, and Follow-up“, *Wounds : a compendium of clinical research and practice*, Jg. 29, S19–S36, 2017.
- [19] K. Reimers, M. Aust und H. Sorg, „Biologische Grundlagen der Wundheilung“, in *Praxis der Plastischen Chirurgie: Plastisch-rekonstruktive Operationen Plastisch-ästhetische Operationen Handchirurgie Verbrennungs chirurgie*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, S. 3–11, ISBN: 978-3-540-37573-9. DOI: 10.1007/978-3-540-37573-9\_1. Adresse: [https://doi.org/10.1007/978-3-540-37573-9%7B%5C\\_%7D1](https://doi.org/10.1007/978-3-540-37573-9%7B%5C_%7D1).
- [20] C. Hafner, S. Meyer und T. Vogt, „Mechanismen der epithelialen Regeneration“, *Der Hautarzt*, Jg. 53, Nr. 8, S. 561–575, 2002, ISSN: 1432-1173. DOI: 10.1007/s00105-002-0412-2. Adresse: <https://doi.org/10.1007/s00105-002-0412-2>.
- [21] G. Han und R. Ceilley, „Chronic wound healing: a review of current management and treatments“, *Advances in therapy*, Jg. 34, Nr. 3, S. 599–610, 2017.
- [22] Medetec, *Medetec*. Adresse: <https://medetec.co.uk/index.html> (besucht am 12.04.2020).
- [23] W. Scherbaum, *Diabetes-Deutschland*, 2009. Adresse: <https://www.diabetes-deutschland.de/> (besucht am 10.06.2020).

- [24] M. E. Edmonds und A. V. M. Foster, „Diabetic foot ulcers“, eng, *BMJ (Clinical research ed.)*, Jg. 332, Nr. 7538, S. 407–410, Feb. 2006, ISSN: 1756-1833. doi: 10.1136/bmj.332.7538.407. Adresse: [https://pubmed.ncbi.nlm.nih.gov/16484268/](https://pubmed.ncbi.nlm.nih.gov/16484268;); <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1370976/>.
- [25] A. Adeyi, S. Muizerengi, I. G. A. Adeyi und I. Gupta, „Leg ulcers in older people: a review of management leg ulcers in older people: a review of management“, *British Journal of medical practitioners*, Jg. 2, Nr. 3, 2009.
- [26] R. L. Hardman, O. Jazaeri, J. Yi, M. Smith und R. Gupta, „Overview of classification systems in peripheral artery disease“, eng, *Seminars in interventional radiology*, Jg. 31, Nr. 4, S. 378–388, Dez. 2014, ISSN: 0739-9529. doi: 10.1055/s-0034-1393976. Adresse: <https://pubmed.ncbi.nlm.nih.gov/25435665/>; <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4232437/>.
- [27] B. Vasudevan, „Venous leg ulcers: Pathophysiology and Classification“, eng, *Indian dermatology online journal*, Jg. 5, Nr. 3, S. 366–370, Juli 2014, ISSN: 2229-5178. doi: 10.4103/2229-5178.137819. Adresse: <https://pubmed.ncbi.nlm.nih.gov/25165676/>; <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4144244/>.
- [28] M. Whiteley, *CEAP clinical score*. Adresse: <https://thewhiteleyclinic.co.uk/self-assessment/ceap-classification/> (besucht am 15.06.2020).
- [29] J. E. Grey, K. G. Harding und S. Enoch, „Venous and arterial leg ulcers“, eng, *BMJ (Clinical research ed.)*, Jg. 332, Nr. 7537, S. 347–350, Feb. 2006, ISSN: 1756-1833. doi: 10.1136/bmj.332.7537.347. Adresse: <https://pubmed.ncbi.nlm.nih.gov/16470058/>; <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1363917/>.
- [30] O. Simeone, „A Very Brief Introduction to Machine Learning With Applications to Communication Systems“, *IEEE Transactions on Cognitive Communications and Networking*, Jg. 4, Nr. 4, S. 648–664, 2018.
- [31] S. Shalev-Shwartz und S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.
- [32] S. Russell und P. Norvig, „Artificial intelligence: a modern approach 3rd Edition“, S. 694–695, 2016.
- [33] ——, „Artificial intelligence: a modern approach 3rd Edition“, S. 830–831, 2016.
- [34] I. Krasin, T. Duerig, N. Alldrin, V. Ferrari, S. Abu-El-Haija, A. Kuznetsova, H. Rom, J. Uijlings, S. Popov, S. Kamali, M. Mallochi, J. Pont-Tuset, A. Veit, S. Belongie, V. Gomes, A. Gupta, C. Sun, G. Chechik, D. Cai, Z. Feng, D. Narayanan und K. Murphy, „OpenImages: A public dataset for large-scale multi-label and multi-class image classification.“, *Dataset available from https://storage.googleapis.com/openimages/web/index.html*, 2017.
- [35] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár und C. L. Zitnick, „Microsoft COCO: Common Objects in Context“, in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele und T. Tuytelaars, Hrsg., Cham: Springer International Publishing, 2014, S. 740–755, ISBN: 978-3-319-10602-1. Adresse: <http://cocodataset.org/#home>.

- [36] Google, *Tensorflow - Transfer learning*, 2020. Adresse: [https://www.tensorflow.org/tutorials/images/transfer\\_learning](https://www.tensorflow.org/tutorials/images/transfer_learning); <https://developers.google.com/machine-learning/glossary#t>; [https://www.tensorflow.org/js/tutorials/transfer/what\\_is\\_transfer\\_learning](https://www.tensorflow.org/js/tutorials/transfer/what_is_transfer_learning) (besucht am 06.06.2020).
- [37] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid und S. Savarese, „Generalized Intersection over Union“, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Juni 2019.
- [38] G. Developers, *Machine Learning Crash Course: Classification*, 2020. Adresse: <https://developers.google.com/machine-learning/crash-course/classification/>.
- [39] L. Derczynski, „Complementarity, F-score, and NLP Evaluation“, in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016, S. 261–266.
- [40] L. Fei-Fei, K. Ranjay und D. Xu, *Lecture - Convolutional Neural Networks for Visual Recognition*, 2020. Adresse: <http://cs231n.stanford.edu>.
- [41] ———, *Lecture - Convolutional Neural Networks for Visual Recognition, Module 1.5*, 2020. Adresse: <https://cs231n.github.io/neural-networks-1/>.
- [42] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006, S. 256–272.
- [43] Chabacano, *Overfitted function*, 2008. Adresse: <https://en.wikipedia.org/wiki/Overfitting#/media/File:Overfitting.svg>.
- [44] A. Krizhevsky, I. Sutskever und G. E. Hinton, „ImageNet Classification with Deep Convolutional Neural Networks“, in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou und K. Q. Weinberger, Hrsg., Curran Associates, Inc., 2012, S. 1097–1105. Adresse: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.
- [45] L. Fei-Fei, K. Ranjay und D. Xu, *Lecture - Convolutional Neural Networks for Visual Recognition, Module 2.1*, 2020. Adresse: <https://cs231n.github.io/convolutional-networks/>.
- [46] R. Girshick, „Fast R-CNN“, in *The IEEE International Conference on Computer Vision (ICCV)*, Dez. 2015.
- [47] R. Girshick, J. Donahue, T. Darrell und J. Malik, „Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation“, in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Juni 2014.
- [48] S. Ren, K. He, R. Girshick und J. Sun, „Faster r-cnn: Towards real-time object detection with region proposal networks“, in *Advances in neural information processing systems*, 2015, S. 91–99.
- [49] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers und A. W. M. Smeulders, „Selective search for object recognition“, *International journal of computer vision*, Jg. 104, Nr. 2, S. 154–171, 2013.

- [50] M. D. Zeiler und R. Fergus, „Visualizing and Understanding Convolutional Networks“, in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele und T. Tuytelaars, Hrsg., Cham: Springer International Publishing, 2014, S. 818–833, ISBN: 978-3-319-10590-1.
- [51] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006, S. 115–116.
- [52] M. Goyal, N. D. Reeves, A. K. Davison, S. Rajbhandari, J. Spragg und M. H. Yap, „DFUNet: Convolutional Neural Networks for Diabetic Foot Ulcer Classification“, *IEEE Transactions on Emerging Topics in Computational Intelligence*, S. 1–12, 2018.
- [53] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke und A. Rabinovich, „Going deeper with convolutions“, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, S. 1–9.
- [54] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard und L. D. Jackel, „Backpropagation Applied to Handwritten Zip Code Recognition“, *Neural Computation*, Jg. 1, Nr. 4, S. 541–551, 1989. doi: 10.1162/neco.1989.1.4.541. Adresse: <https://doi.org/10.1162/neco.1989.1.4.541>.
- [55] X. Zhao, Z. Liu, E. Agu, A. Wagh, S. Jain, C. Lindsay, B. Tulu, D. Strong und J. Kan, „Fine-Grained Diabetic Wound Depth and Granulation Tissue Amount Assessment Using Bilinear Convolutional Neural Network“, *IEEE Access*, Jg. 7, S. 179151–179162, 2019, ISSN: 2169-3536. doi: 10.1109/ACCESS.2019.2959027.
- [56] P. E. Houghton, C. B. Kincaid, K. E. Campbell, M. G. Woodbury und D. H. Keast, „Photographic assessment of the appearance of chronic pressure and leg ulcers“, *Ostomy Wound Management*, Jg. 46, Nr. 4, S. 20–35, 2000.
- [57] K. Simonyan und A. Zisserman, „Very deep convolutional networks for large-scale image recognition“, *arXiv preprint arXiv:1409.1556*, 2014.
- [58] J. Deng, W. Dong, R. Socher, L. Li, Kai Li und Li Fei-Fei, „ImageNet: A large-scale hierarchical image database“, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Juni 2009, S. 248–255. doi: 10.1109/CVPR.2009.5206848.
- [59] T. W. K. Poon und M. R. Friesen, „Algorithms for Size and Color Detection of Smartphone Images of Chronic Wounds for Healthcare Applications“, *IEEE Access*, Jg. 3, S. 1799–1808, 2015, ISSN: 2169-3536. doi: 10.1109/ACCESS.2015.2487859.
- [60] T. Biswas, M. F. Ahmad Fauzi, F. S. Abas und H. K. R. Nair, „Superpixel Classification with Color and Texture Features for Automated Wound Area Segmentation“, in *2018 IEEE Student Conference on Research and Development (SCOReD)*, Nov. 2018, S. 1–6. doi: 10.1109/SCOReD.2018.8711241.
- [61] C. Wang, X. Yan, M. Smith, K. Kochhar, M. Rubin, S. M. Warren, J. Wrobel und H. Lee, „A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks“, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, S. 2415–2418.

- [62] A. Khalil, M. Elmogy, M. Ghazal, C. Burns und A. El-Baz, „Chronic Wound Healing Assessment System Based on Different Features Modalities and Non-Negative Matrix Factorization (NMF) Feature Reduction“, *IEEE Access*, Jg. 7, S. 80110–80121, 2019, ISSN: 2169-3536. doi: 10.1109/ACCESS.2019.2923962.
- [63] A. Yee, F. Meng, J. Harmon und S. Yi, „An Innovative App for the Management of Chronic Wound Treatment“, in *Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies*, Ser. CHASE ’18, New York, NY, USA: Association for Computing Machinery, 2018, S. 3–4, ISBN: 9781450359580. doi: 10.1145/3278576.3281273. Adresse: <https://doi-org.eaccess.ub.tum.de/10.1145/3278576.3281273>.
- [64] S. Wang, Q. Zhang, W. Huang, H. Tian, J. Hu, Y. Cheng und Y. Peng, „A New Smart Mobile System for Chronic Wound Care Management“, *IEEE Access*, Jg. 6, S. 52355–52365, 2018, ISSN: 2169-3536. doi: 10.1109/ACCESS.2018.2864264.
- [65] K. Klinker, M. Wiesche und H. Krcmar, „Development of a Smart Glass Application for Wound Management“, in *Extending the Boundaries of Design Science Theory and Practice*, B. Tulu, S. Djamasbi und G. Leroy, Hrsg., Cham: Springer International Publishing, 2019, S. 157–171, ISBN: 978-3-030-19504-5.
- [66] ——, „Digital transformation in health care: Augmented reality for hands-free service innovation“, *Information Systems Frontiers*, S. 1–13, 2019.
- [67] T. Fischer, A. Baczako, A. Konstantinow und T. Volz, „Chronische Wunden richtig behandeln“, *hautnah dermatologie*, Jg. 35, Nr. 5, S. 44–51, 2019.
- [68] H. Wüller, J. Behrens, K. Klinker, M. Wiesche, H. Krcmar und H. Remmers, „Smart Glasses in Nursing-Situation Change and Further Usages Exemplified on a Wound Care Application.“, in *GMDS*, 2018, S. 191–195.
- [69] M. Maity, D. Dhane, C. Bar, C. Chakraborty und J. Chatterjee, „Pixel-Based Supervised Tissue Classification of Chronic Wound Images with Deep Autoencoder“, in *Advanced Computational and Communication Paradigms*, S. Bhattacharyya, N. Chaki, D. Konar, U. K. Chakraborty und C. T. Singh, Hrsg., Singapore: Springer Singapore, 2018, S. 727–735, ISBN: 978-981-10-8237-5.
- [70] E. Electronics, *How To Train an Object Detection Classifier Using TensorFlow (GPU) on Windows 10*, 2018. Adresse: <https://www.youtube.com/watch?v=Rgpfk6eYxJA%7B%5C%7Dt=130s>.
- [71] L. M. Hohmann, Google und E. Electronics, *Code-Repository*, 2020. Adresse: <https://gitlab.lrz.de/ga27can/bachelorarbeit-code>.
- [72] S. Ltd., *Affinity Photo*. Adresse: <https://affinity.serif.com/de/photo/> (besucht am 07.04.2020).
- [73] P. Reindl-Spanner, „Automatische Wunderkennung durch Verwendung von Machine Learning“, IDP, TUM, 2020.

- [74] W. Reinhardt, E. Ruegenhagen und B. Rummel, „Fragebogen zur System-Gebrauchstauglichkeit“, Techn. Ber., 2013. Adresse: <https://experience.sap.com/skillup/system-usability-scale-jetzt-auch-auf-deutsch/>.
- [75] J. Brooke u. a., „Sus-a quick and dirty usability scale“, *Usability evaluation in industry*, Jg. 189, Nr. 194, S. 4–7, 1996.
- [76] Usability.gov, *System Usability Scale (SUS)*. Adresse: <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html> (besucht am 30.05.2020).
- [77] Medetec und L. M. Hohmann, *Bilddaten*, 2020. Adresse: <https://www.dropbox.com/sh/oaoon0uf4p2v04mw/AABXUkmRnAbPkQp5KEFkf50Sa?dl=0>.
- [78] K. Anthony und E. Luo, *Arterial and Venous Ulcers: What's the Difference?*, 2018. Adresse: <https://www.healthline.com/health/arterial-vs-venous-ulcers> (besucht am 07.05.2020).
- [79] A. Gabriel, M. Camp, C. Paletta und B. Massey, *Vascular Ulcers*, 2018. Adresse: <https://emedicine.medscape.com/article/1298345-overview#showall> (besucht am 07.05.2020).
- [80] L. Loretz, *Arterial and Venous Lower Extremity Ulcers*. Adresse: [https://medicine.wright.edu/sites/medicine.wright.edu/files/page/attachments/LLoretzArterialVenousLowerExtUlcers-final%5B2%5D\\_0.pdf](https://medicine.wright.edu/sites/medicine.wright.edu/files/page/attachments/LLoretzArterialVenousLowerExtUlcers-final%5B2%5D_0.pdf) (besucht am 07.05.2020).
- [81] Google, *Descending into ML: Training and Loss*, 2020. Adresse: <https://developers.google.com/machine-learning/crash-course/descending-into-ml/training-and-loss> (besucht am 20.06.2020).
- [82] F. Bashir und F. Porikli, „Performance evaluation of object detection and tracking systems“, in *Proceedings 9th IEEE International Workshop on PETS*, 2006, S. 7–14.
- [83] U. Cambridge, *Introduction to information retrieval*. 2009, S. 162. Adresse: <https://dspace.echhost.com/jspui/bitstream/123456789/2452/1/00776216.pdf>.
- [84] M. Everingham und J. Winn, *The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Development Kit*. Adresse: [http://host.robots.ox.ac.uk/pascal/VOC/voc2012/html/doc/devkit\\_doc.html](http://host.robots.ox.ac.uk/pascal/VOC/voc2012/html/doc/devkit_doc.html) (besucht am 23.06.2020).
- [85] A. B. Jung, K. Wada, J. Crall, S. Tanaka, J. Graving, C. Reinders, S. Yadav, J. Banerjee, G. Vecsei, A. Kraft, Z. Rui, J. Borovec, C. Vallentin, S. Zhydenko, K. Pfeiffer, B. Cook, I. Fernández, F.-M. De Rainville, C.-H. Weng, A. Ayala-Acevedo, R. Meudec, M. Laporte u. a., *ImgAug*, 2020. Adresse: <https://github.com/aleju/imgaug/>.
- [86] B. Zoph, E. D. Cubuk, G. Ghiasi, T.-Y. Lin, J. Shlens und Q. V. Le, „Learning data augmentation strategies for object detection“, *arXiv preprint arXiv:1906.11172*, 2019.
- [87] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006, S. 059.